Journal of Psychiatric Research 50 (2014) 73–78

Contents lists available at ScienceDirect

Journal of Psychiatric Research

journal homepage: www.elsevier.com/locate/psychires

Protein-interaction-network-based analysis for genome-wide association analysis of schizophrenia in Han Chinese population

Hao Yu [a,b,1], Wenjian Bi [c,1], Chenxing Liu [a,b], Yanlong Zhao [c], Ji-Feng Zhang [c], Dai Zhang [a,b,d,e,*], Weihua Yue [a,b,*]

[a] Institute of Mental Health, Peking University, 51 Hua Yuan Bei Road, Beijing 100191, China
[b] Key Laboratory of Mental Health, Ministry of Health, Institute of Mental Health, The Sixth Hospital, Peking University, China
[c] Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China
[d] Peking-Tsinghua Center for Life Sciences, Beijing, PR China
[e] PKU-IDG/McGovern Institute for Brain Research, Peking University, Beijing, China

ARTICLE INFO

Article history:
Received 25 June 2013
Received in revised form
10 November 2013
Accepted 27 November 2013

Keywords:
Schizophrenia
Genome-wide association study
Protein-interaction-network-based analysis
Susceptibility gene
Pathway
Network

ABSTRACT

Schizophrenia is a severe neuropsychiatric disorder with a strong and complex genetic background. Recent genome-wide association studies (GWAS) have successfully identified several susceptibility loci of schizophrenia. In order to interpret the functional role of the genetic variants and detect the combined effects of some of these genes on schizophrenia, protein-interaction-network-based analysis (PINBA) has emerged as an effective approach. In the current study, we conducted a PINBA of our previous GWAS data taken from the Han Chinese population. In order to do so, we used dense module search (DMS), a method that locates densely connected modules for complex diseases by integrating the association signal from GWAS datasets into the human protein–protein interaction (PPI) network. As a result, we identified one gene set with a joint effect significantly associated with schizophrenia and gene expression profiling analysis suggested that they were mainly neuro- and immune-related genes, such as glutamatergic gene (*GRM5*), GABAergic genes (*GABRB1*, *GABARAP*) and genes located in the MHC region (*HLA-C*, *TAP2*, *HIST1H1B*). Further pathway enrichment analysis suggested that these genes are involved in processes related to neuronal and immune systems, such as the Adherens junction pathway, the Neurotrophin signaling pathway and the Toll-like receptor signaling pathway. In our study, we identified a set of susceptibility genes that had been missed in single-marker GWAS, and our findings could promote the study of the genetic mechanisms in schizophrenia.

1. Introduction

Schizophrenia is highly heritable and complex psychiatric disorder with a lifetime prevalence of ~1% and estimated heritability of ~64–80% (Lichtenstein et al., 2009; Thaker & Carpenter, 2001). In recent years, researchers have used genome-wide association studies (GWAS) to successfully identify several susceptibility loci for the disease (O'Donovan et al., 2008; Purcell et al., 2009; Shi et al., 2009; Yue et al., 2011). These GWAS analyses focused on finding the strongest single-nucleotide polymorphisms (SNPs) that met the genome-wide significance cutoff *P*-value of 5×10^{-8} for detecting significant markers. The traditional GWAS typically investigate the genetic effect of a single SNP at a time and it account for only a small proportion of the heritability of schizophrenia, leaving a large portion of the disease's susceptibility unexplained (Eichler et al., 2010; Manolio et al., 2009). Furthermore, schizophrenia is believed to be a multigenic disorder that involves many genes functioning at various stages of disease development. Due to its complex genetic architecture and joint effects among these genes, the overall effect of a gene network is expected to have a greater effect than the sum of individual effect of each gene. Therefore, pathway- and network-based methods have been developed to provide functional links to bridge the knowledge gap between the genetic variants and the phenotypes. Combining with the results from GWAS, these approaches can assess whether a group of genes or pathways with related functions are jointly associated with a trait of interest and generate specific hypothesis for follow-up experimental studies (Sun, 2012).

* Corresponding authors. Institute of Mental Health, Peking University, 51 Hua Yuan Bei Road, Beijing 100191, China. Tel: +86 10 8280 5307; fax: +86 10 6201 7114.
E-mail addresses: daizhang@bjmu.edu.cn (D. Zhang), dryue@bjmu.edu.cn (W. Yue).
[1] Yu H and Bi WJ contributed to this work equally.

0022-3956/$ – see front matter
http://dx.doi.org/10.1016/j.jpsychires.2013.11.014

Compared to pathway-based analysis (PBA), network-based analysis (NBA) of GWAS data has advantages in the following aspects (Jia et al., 2012). First, PBA see the whole pathway as a single unit, however, the association signals from GWAS might cover only a small portion of the pathway reducing its power. Unlike PBA, NBA searches dynamic gene sets, thus relieving the limitation of fixed size in a pathway. Second, the definition of canonical pathway is incomplete and the genes in the pathway cover only a small portion of genes from the GWAS data. For example, the KEGG database covered 5000–5500 genes (Kanehisa, Goto, Furumichi, Tanabe, & Hirakawa, 2010). However, a recent analysis of protein–protein interaction (PPI) data from multiple sources has reconstructed the human PPI network by recruiting ~12,000 proteins and ~60,000 protein interaction pairs (Jia, Zheng, Long, Zheng, & Zhao, 2011).

Protein-interaction-network-based analysis (PINBA) of GWAS data is a recently developed network-based method for identifying susceptibility genes that investigates whether a set of genes with related function is jointly associated with a trait or disease. Instead of focusing on whether or not SNPs are individually significant, PINBA combines GWAS results with prior biological knowledge about protein-interaction to assess associability. This approach may generate new susceptibility genes and provide novel hypotheses for follow-up experiments. PINBA has previously been applied to the research of the underlying biological mechanisms involved in complex diseases, successfully yielding the relevant networks and new susceptibility genes (Baranzini et al., 2009; Jia et al., 2011; Lu et al., 2013). And we intend to find novel susceptibility variants or genes for schizophrenia, PINBA could find susceptibility genes based on SNP-level *P*-value. Presently, no supportive PINBA findings of schizophrenia have been reported in the Chinese Han population.

In the current study, we performed a PINBA on our own previously collected GWAS data (Yue et al., 2011) in order to identify some underlying genetic factors of schizophrenia in the Han Chinese population. First, we performed a PINBA using the Dense Module Searching (DMS) method (Jia et al., 2011), which searched for and assessed dense modules involved in disease pathophysiology by incorporating GWAS datasets into the PPI network. Following the PINBA, we conducted gene expression profiling analysis and pathway enrichment tests of the module genes that we identified in search of a better understanding of the underlying biological processes of schizophrenia. We found that the resultant genes are mainly neural- and immune-related and more likely to interact and take part in the same or related pathways. Of note, additional susceptibility genes were proposed through this approach.

2. Materials and methods

2.1. GWAS data and protein–protein interaction (PPI) datasets

We used the GWAS data from a study we previously conducted (Yue et al., 2011). Our GWAS samples (768 schizophrenia cases and 1733 normal controls) came from individuals of Han Chinese ancestry, genotyped with Illumina Human610-Quad BeadChips. In quality control, we examined potential genetic association based on pairwise identity-by-state analysis for all of the successfully genotyped samples. Upon identification of any probable first- or second-degree relatives pair, we removed one of the two likely related individuals (whichever subject had the lower call rate). One schizophrenia case and two controls were removed because of either missing genotype rates greater than 0.1 or relative relationship with another subject. After quality control, we excluded SNPs with call rates less than 90%, minor allele frequencies less than 5%, and Hardy-Weinberg equilibrium *P*-value less than 1×10^{-5} in the controls. After quality control filtering, a total of 448,734 autosomal SNPs in 746 schizophrenia cases and 1599 normal controls were retained for PINBA.

The study was approved by the Medical Research Ethics Committee of the Institute of Mental Health, Peking University. All participants were given detailed verbal and written information regarding the purpose and procedures of the study. Written consents were obtained from the patients and/or their parents, and all healthy participants enrolled in this study.

We downloaded PPI datasets from the Protein Interaction Network Analysis (PINA) platform (http://cbg.garvan.unsw.edu.au/pina/). To ensure the reliability of the PPI data, we included only those interactions with experimental evidence proving that they took place between human genes. The final network included a total of 11,996 distinct proteins and 72,506 interaction sets.

2.2. PINBA

In the current study, we used a PINBA approach proposed by Jia et al. (2011), in which the dense module search (DMS) method is conducted in an R package that they developed, called '*dmGWAS*' (http://bioinfo.mc.vanderbilt.edu/dmGWAS.html). Our analysis consisted of three main steps.

1) First, a SNP from the GWAS data was mapped to a gene if its position was within that gene's National Center for Biotechnology Information (NCBI) annotated start and stop coordinates. In order to account for variants in potential gene control regions, we also included SNPs located 20 kb upstream and 20 kb downstream of each gene. We selected the most significant SNP, whose *P*-value was smallest among the SNPs within a gene, to represent the extent of association of gene with the schizophrenia.
2) Using the DMS method, we searched for the subnetwork, or module, that owned a maximum proportion of low *P*-value genes within the whole human PPI network. A score (Z_m) was computed using the following formula:

$$Z_m = \sum Z_i/\sqrt{w} \quad i = (1, 2, 3 \ldots w)$$

w represented the number of genes within a module. Z_i was computed using the formula: $Z_i = \phi^{-1}(1 - P_i)$, where ϕ^{-1} represented the inverse normal distribution function (Ideker, Ozier, Schwikowski, & Siegel, 2002) and P_i represented the *P*-value of a gene. Then, we performed a searching strategy with different parameters in the *dmGWAS* software (i.e. d and r). The parameter d represented a predefined distance constraint and r was the rate of proportion increment, nodes will be added if the increment is greater than $Z_m \times r$. In previous study, based on the fact that the median distance between any two proteins in the human PPI network is less than 5 and parameter d has a marginal effect on the results, it's recommended that d is set as the default value 2. However, the parameter r has a substantial effect on the results. When r is small, it applies a loose restriction during the module expanding process; thus, unrelated nodes might be included. On the other hand, when r is large, a strict restriction is imposed and only those nodes with very high Z_i value could be included (Jia et al., 2011). Therefore, we compared the resultant modules generated under parameter $d = 2$ and different r values (0.05, 0.1, 0.15 and 0.2) and chose the appropriate parameters for later analysis. Finally, we performed DMS with the appropriate parameters.

3) To assess the significance of the identified modules, we built two distributions under two hypotheses (Jia et al., 2011). The first