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

null hypothesis was that there is no difference between the identified modules and modules randomly selected from the whole network (Jia et al., 2011). If a module had n genes, we randomly chose the same number of genes from the whole network, computed Z_m, and denoted it by Z_m (k). To achieve sufficient randomization, this process was repeated 100,000 times. A score (Z_N) was computed with the following formula:

$$Z_N \frac{Z_m - \text{mean}(Z_m(k))}{SD(Z_m)}$$

Accordingly, Z_N was adjusted for gene size and used to compare different modules. Furthermore, Z_N could be transformed back to P-value (P_{Z_N}).

The second null hypothesis was that there is no association between the modules and schizophrenia (Jia et al., 2011). We permutated the disease labels of all samples using PLINK (Purcell et al., 2007), while maintaining the same case/control ratio. During each permutation (denoted by π), we repeated the calculation of Z_m as described above, except that the disease phenotypes were obtained from permutation. We denoted the corresponding Z_m score by Z_m (π). A empirical P-value (P_{per}) was then computed for each module by counting the number of permutations that have Z_m (π) greater than the real case, divided by the total number of permutations ($N = 1000$). Modules with $P_{\text{per}} < 0.01$ were considered statistically significant. The simple two-step correction procedure effectively adjusted for different sizes of genes and preserved the type I error rate below a certain threshold (Jia et al., 2011; Wang, Li, & Bucan, 2007).

After identification of the top-ranking modules, we searched the literature to see if the resultant genes were known as human schizophrenia candidate genes. We used the GWAS catalog developed by Hindorff et al. (2009). We also compared the resultant genes to the work of Ayalew et al. (2012), because they integrated many pivotal datasets from schizophrenia studies, the genes identified represent high confidence candidate genes for schizophrenia.

2.3. Gene expression profiling analysis and pathway enrichment test

To assess cell-specific expression of the identified genes in the modules, we assessed the genes with an online tool Gene Enrichment Profiler (http://xavierlab2.mgh.harvard.edu/EnrichmentProfiler/). This profiler computes the expression and enrichment of any set of query genes on the basis of a reference set obtained from 126 normal tissues and cell types (represented by 557 microarrays). For the enrichment analyses, the 'Gene Enrichment Profiler' calculated the enrichment score using the Linear Models for Microarray Data (LIMMA) module in bioconductor. Given a large body atlas size data set, each tissue is compared pairwise to each of the other tissues and then a linear model coefficient for each pairwise comparison was computed. That coefficient is a measure of difference between two groups. Significant coefficient is P-value less than 0.05. The enrichment score is the sum of all pairwise comparisons for each gene. Thus, the enrichment score can be used to benchmark expression levels in one tissue compared to all other tissues to identify genes that are tissue specific. The widely expressed genes score low and around 0, while genes that are specific score closer to 1. Moreover, in the transcript expression heatmap, colors correspond to the level of enrichment in each tissue or cell type (green = depleted, black = no enrichment, red = enriched). Further information about this online tool can be found in the work completed by Benita et al. (2010).

In addition, to explore the biological significance of the identified genes, we performed pathway enrichment analysis by using DAVID Bioinformatics Resources 6.7 (Huang da, Sherman, & Lempicki, 2009). Fisher exact tests were conducted in DAVID to compute the P-value for each KEGG pathway (Kanehisa et al., 2010), which were then adjusted by Benjamini & Hochberg (BH)'s method (Benjamini & Hochberg, 1995). Since the adjusted P-value <0.05, the query gene list is specifically enriched in the resultant pathway than random chance.

In addition to the running time of PLINK, within a Windows framework, it took ~12 h on a server (3.00 GHz Quad Core Intel (R) CPU Q9650 and 4.00 GB of RAM, four threads) to perform our PINBA.

3. Results

3.1. PINBA

Using different parameters (d, r) in *dmGWAS*, we obtained gene sets with different module sizes. Fig. 1 showed the comparison of module size with different r values. When r is small, the size of modules tends to be large (e.g., when $r = 0.05$, average number of module genes is 17.7), and non-specific nodes might be included and, thus, weaken the signal. When r is large, the size of the modules becomes small (e.g., when $r = 0.2$, it is 6.7, slightly larger than the minimum number of genes for a module, 5), and possibly excluding informative genes from the module. It illustrated that our results using $d = 2$ and $r = 0.1$ worked well and sensitively (average number of genes is 10.6). Then, we performed DMS with the appropriate parameters ($d = 2$, $r = 0.1$) to search candidate modules that were significantly enriched for schizophrenia. We provide the information of gene-wise P-value in Table S1, which could subsequently be used to calculate Z_m in this study.

After performing dense module search, we observed that there are 8472 candidate modules. The normalized module score (Z_N) of this module set was within the range of 3.24 to 7.70. Using the formula: $P_{Z_N} = 1 - \phi(Z_N)$, we transformed the Z_N back to P-value. However, P_{Z_N} was within the range of 5.98×10^{-4} to 6.77×10^{-5},

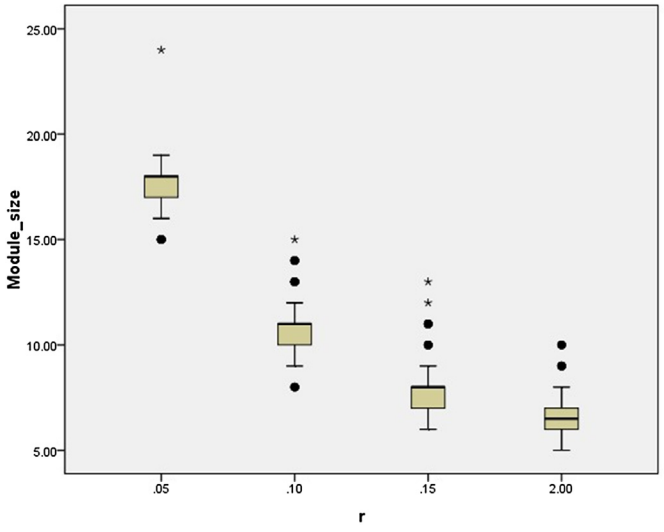

Fig. 1. Comparison of module size with different r values in dense module searching. It showed that when r is small, the size of modules tends to be large (e.g., when $r = 0.05$, average number of module genes is 17.7). When r is large, the size of the modules becomes small (e.g., when $r = 0.2$, it is 6.7, slightly larger than the minimum number of genes for a module, 5). It illustrated that our results using $d = 2$ and $r = 0.1$ worked well and sensitively (average number of genes is 10.6).

Table 1
Pathway enrichment analysis of module genes.

Pathway names	Counts[a]	Pathway size[b]	*P*-value[c]	P_{adj}[d]	Genes[e]
hsa04520:Adherens junction	9	73	1.43E-05	1.35E-03	*MAP3K7, PTPRJ, TJP1, RAC1, SMAD3, ACTN1, SMAD2, PTPN1, SRC*
hsa05212:Pancreatic cancer	7	70	6.08E-04	2.85E-02	*RAC1, SMAD3, RALA, PIK3CA.NFKB1, SMAD2, STAT3*
hsa05200:Pathways in cancer	8	108	8.69E-04	2.72E-02	*PRKCA, GRB2, SMAD3, RUNX1T1, FADD, RAC1, NFKB1, SMAD2, ZBTB16, STAT3, CBLC, RASSF5, RALA, PIK3CA*
hsa04660:T cell receptor signaling pathway	8	127	9.71E-04	2.28E-02	*MAP3K7, CBLC, GRB2, CARD11, PAK7, , PIK3CA, NFKB1, MAP2K7*
hsa05221:Acute myeloid leukemia	8	132	1.52E-03	2.85E-02	*GRB2, RUNX1T1, PIK3CA, NFKB1, ZBTB16, STAT3*
hsa04012:ErbB signaling pathway	7	87	1.65E-03	2.58E-02	*PRKCA, CBLC, PAK7, GRB2, PIK3CA, MAP2K7, SRC*
hsa04722:Neurotrophin signaling pathway	8	128	2.17E-03	2.91E-02	*GRB2, RAC1, YWHAQ, PIK3CA, NFKB1, KIDINS220, MAP2K7, PTPN11*
hsa04530:Tight junction	6	57	3.37E-03	3.93E-02	*PRKCA, EPB41L2, INADL, TJP1, CLDN1, PRKCH, ACTN1, SRC*
hsa04620:Toll-like receptor signaling pathway	7	102	3.52E-03	3.65E-02	*MAP3K7, RAC1, PIK3CA, NFKB1, FADD, IL12B, MAP2K7*
hsa04662:B cell receptor signaling pathway	8	155	4.72E-03	4.39E-02	*CARD11, FCGR2B, GRB2, RAC1, PIK3CA, NFKB1*

[a] The total number of module genes observed in the pathway.
[b] The total number of genes in this pathway.
[c] *P*-values from pathway enrichment test.
[d] *P*-values adjusted by Benjamini & Hochberg (BH) method (Benjamini & Hochberg, 1995), the significance cutoff *P*-value 0.05 was used.
[e] The list of identified genes in the pathway.

thus, it would generate too many modules with P_{Z_N} as the criterion, so we selected the modules whose Z_N scores were within the top 1% in module score distribution for the follow up analysis (Jia et al., 2011). After the permutation tests, all the selected modules were significant ($P_{\text{per}} < 1 \times 10^{-3}$), indicating that they were not only significantly enriched but also associated with schizophrenia.

We listed the gene-wise *P*-values of 184 resultant genes at the top of Table S1. To better understand the interaction among the resultant genes networks, we listed the protein-interaction involved in the 184 resultant genes in Table S2.

From GWAS catalog, we found that *VRK2*, *NFKB1*, *SPTLC1*, *ERC2*, *SLC17A1* had all been previously identified in GWAS (Bergen et al., 2012; Curtis et al., 2011; Liou et al., 2012; Shi et al., 2009; Steinberg et al., 2011). It is worth noting that *NFKB1* was also identified in GWAS performed in the Han Chinese population (Liou et al., 2012). Compared to the resultant genes to the work of Ayalew et al. (2012), we found 6 overlapped genes (*DISC1*, *GRIN2B*, *GSN*, *PRKCA*, *PDE4B*, and *GRM5*). These demonstrates that our results confirmed previous findings.

3.2. Gene expression profiling analysis and pathway enrichment analysis

In the transcript expression heatmap (Fig. S1), more than half of these genes were highly expressed in immune-related cell types (specifically B, T and myeloid cells) and the CNS tissues. In addition, from transcript enrichment heatmap (Fig. S2), we found genes identified in the PPI network were preferentially expressed in CNS tissues, immune cell types and tissues. Particularly, the results of our pathway enrichment analysis also suggested that these genes are involved in processes related to neural and immune systems. Table 1 shows the results of the pathway enrichment analyses of the identified gene set by DAVID using KEGG pathways (Kanehisa et al., 2010).

4. Discussion

Schizophrenia is a complex neuropsychiatric disorder with high heritability. To date, multiple GWAS of schizophrenia have been conducted and many pivotal schizophrenia susceptibility genes have been identified. So far, most traditional GWAS have focused on the strongest association signals, but many weakly or moderately significant signals may also provide a valuable insight. A larger number of loci with weakly or moderately significant signals may affect relevant mechanisms in a cumulative effect. Therefore, in this study, we have applied DMS to identify densely connected modules in the human protein-interaction that conjoins loci of strong to weak significance. By performing the DMS, we successfully identified a gene set of 184 susceptibility genes that may play a role in the etiology of schizophrenia. Most of the resultant genes had relatively moderate effect sizes, making them unlikely findings in traditional single-marker Genome-Wide-Analyses (GWA) of schizophrenia.

The module genes that we identified mainly consisted of neural- and immune-related genes, for example, we found genes involved in GABA receptor (*GABRB1*, *GABARAP*), genes from the 14-3-3 protein family (*YWHAQ*), genes involved in glutamate receptor (*GRM5*), genes located in the MHC region (*HLA-C*, *TAP2*, *HIST1H1B*) and well-studied candidate genes for schizophrenia (*DISC1*, *GRIN2B*). In addition, we also conducted the gene expression profiling analysis, and found most genes identified in the PPI network were preferentially expressed in CNS tissues, immune cell types and tissues. Furthermore, when performing the pathway enrichment analysis, we found the resultant pathways are related with the neuropathology and immune systems. The results of pathway enrichment analysis conform to the findings of Gene Expression Profiling Analysis. The results not only provide further evidence that schizophrenia is a complex disease involving immune systems (Purcell et al., 2009; Shi et al., 2009; Stefansson et al., 2009), but also indicate that the resultant genes identified in the modules are expressed in brain tissues, suggesting these genes may play important roles in brain function. In fact, more and more evidence suggested immune-related genes may play important roles in schizophrenia. For example, several GWAS revealed that many immune genes are significantly associated with schizophrenia (Ripke et al., 2011; Stefansson et al., 2009; Steinberg et al., 2011), and dysregulation of immune-associated genes were frequently observed in schizophrenia patients (Fillman et al., 2013; Gardiner et al., 2013). Altogether, our findings could help form a better understanding of the implications of gene sets on schizophrenia progression and lead to the identification of susceptibility genes associated with these pathways, which could be possible targets for drug development. From these identified gene sets, we might discover better and more accurate biomarkers for personalized therapies and treatment (Barabasi, Gulbahce, & Loscalzo, 2011). For example, in our findings, *GRIN2B* genetic variations might influence the effect of clozapine on Chinese patients (Chiu, Wang, Liou, Lai, & Chen, 2003). *DISC1* was found to be statistically linked with ultra-resistance to antipsychotic treatment (Mouaffak et al., 2011).

Our findings also confirmed the earlier observation that immune-related and neural pathways were involved in schizophrenia susceptibility. O'Dushlaine et al. found several neural pathways involved in processes critical to neurodevelopment and

synaptic function (O'Dushlaine et al., 2011). Jia et al. also identified several immune related pathways associated with schizophrenia (Jia et al., 2012). In our results, the Neurotrophin signaling pathway (hsa04722) plays an important role in differentiation and survival of neural cells and is involved in key stages of the formation of the neuronal network; the Tight junction (hsa04530) pathway is relevant to synaptic formation and neurotransmission at glutamatergic and GABAergic synapses (Kang, Zhang, Dobie, Wu, & Craig, 2008); the ErbB signaling pathway (hsa04012) participates in cognitive dysfunction in schizophrenia by aberrantly suppressing Src-mediated enhancement of synaptic NMDAR function (Pitcher et al., 2011). The T cell receptor signaling pathway (hsa04660), B cell receptor signaling pathway (hsa04662) and the Toll-like receptor signaling pathway (hsa04620) are well-known pathways involved in immune system function. Interestingly, we find four pathways involved in malignant tumors, the pancreatic cancer pathway (hsa05212), the Pathways in cancer (hsa05200), the Acute myeloid leukemia pathway (hsa05221), and the Chronic myeloid leukemia pathway (hsa05220). However, the identified genes in the four pathways are mainly immune-related genes and the details are shown in Table 1.

To our knowledge, our study is the first attempt at integration of protein-interaction data with GWAS data in relation to schizophrenia in Han Chinese population. Our findings not only confirmed previous findings, but also highlighted new susceptibility genes and pathways underlying schizophrenia. These new-found genes and pathways may be peculiar susceptibility genes in Han Chinese population, but further experimental approaches are warranted for determining which are the functionally relevant associations in each of these genes. These resultant genes were not found in our previous GWAS. Hence, the PINBA method has identified some of the missing heritability of our GWAS data and will contribute to the identification of new genetic factors underlying schizophrenia in the Han Chinese population.

Our network-based approach has strengths and limitations. Compared with pathway-based methods, the network-based analysis allows for the definition of *de novo* gene sets by dynamically searching for interconnected subnetworks in the whole PPI network, but the pathway-based analysis over-limited to a priori knowledge. In addition, the association signals from GWAS might cover only a small portion of the pathway reducing its power; however, the network-based methods effectively alleviate the limitation of the fixed size in pathway analysis. This work also has a few limitations. First, it is still not clear what the best strategy is to condense statistics for multiple SNPs within a gene into a single value for the gene (Wang, Li, & Hakonarson, 2010). In present study, we used the most popular method, taking the minimum SNP-level *P*-value to represent the significance of each gene (Arning et al., 2012; Wang et al., 2007; Wang et al., 2010). However, using this strategy, larger genes are more likely to have lower minimum *P*-values based on chance alone. And we have also compared the SNP numbers for each of the 184 resultant genes to all genes in the genome and found there was significant difference (Mann–Whitney test, $P < 0.05$). A partial remedy for such a bias is to normalize gene-set score. And we used a two-step correction procedure and preserved the type I error rate across genes of different size. Moreover, the 184 reported genes have all passed the corrections. Second, our network-based analysis was based on computational strategies. The modules were built on available human PPI datasets. The informative results generated by this approach generally require extensive experimental validation and the number and quality of protein interactions has recently improved greatly, the human PPI datasets are still constantly improving. Third, similar to single-marker-based association test, network-based analysis may also be susceptible to false-positive results and thus should be appropriately replicated in independent datasets. Due to limitation of conditions, we have not validated the association by replication analysis at this point in time. And we intend to verify the results in a replication study in the future.

In conclusion, our approach suggests that integration of protein-interaction data with a GWA study of schizophrenia can yield potentially interesting sets of susceptibility genes that would be missed in traditional GWAS. We identified one gene set of 184 genes that mainly preferentially expressed in CNS tissues, immune cell types and tissues. Additionally, the resultant genes were also enriched in neural- and immune-related pathways. Confirmation of these genes in replication studies or follow-up experimental studies may promote the study of the genetic mechanisms of schizophrenia in Han Chinese population.

Role of the funding source

This work was supported by grants from the National Natural Science Foundation of China (81071087, 81222017, 11171333, 81221002 and 61134013); Program for New Century Excellent Talents in University (NCET); National High Technology Research and Development (2009AA022702); and International Science & Technology Cooperation Program of China (2010DFB30820).

Contributors

Authors WHY and DZ designed the study. HY, WJB and CXL carried out the analyses. YLZ, JFZ and WHY provided guidance on data analysis. HY, WJB and WHY wrote the manuscript.

Conflict of interest

All authors have no conflicts of interest to declare.

Acknowledgements

We extend our gratitude to all participants who participated in this study.

Appendix A. Supplementary material

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.jpsychires.2013.11.014.

References

Arning A, Hiersche M, Witten A, Kurlemann G, Kurnik K, Manner D, et al. A genome-wide association study identifies a gene network of ADAMTS genes in the predisposition to pediatric stroke. Blood 2012;120:5231–6.
Ayalew M, Le-Niculescu H, Levey DF, Jain N, Changala B, Patel SD, et al. Convergent functional genomics of schizophrenia: from comprehensive understanding to genetic risk prediction. Molecular Psychiatry 2012;17:887–905.
Barabasi AL, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. Nature Reviews Genetics 2011;12:56–68.
Baranzini SE, Galwey NW, Wang J, Khankhanian P, Lindberg R, Pelletier D, et al. Pathway and network-based analysis of genome-wide association studies in multiple sclerosis. Human Molecular Genetics 2009;18:2078–90.
Benita Y, Cao Z, Giallourakis C, Li C, Gardet A, Xavier RJ. Gene enrichment profiles reveal T-cell development, differentiation, and lineage-specific transcription factors including ZBTB25 as a novel NF-AT repressor. Blood 2010;115:5376–84.
Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society: Series B 1995;57:289–300.
Bergen SE, O'Dushlaine CT, Ripke S, Lee PH, Ruderfer DM, Akterin S, et al. Genome-wide association study in a Swedish population yields support for greater CNV and MHC involvement in schizophrenia compared with bipolar disorder. Molecular Psychiatry 2012;17:880–6.
Chiu HJ, Wang YC, Liou YJ, Lai IC, Chen JY. Association analysis of the genetic variants of the N-methyl D-aspartate receptor subunit 2b (NR2b) and treatment-refractory schizophrenia in the Chinese. Neuropsychobiology 2003;47:178–81.

Curtis D, Vine AE, McQuillin A, Bass NJ, Pereira A, Kandaswamy R, et al. Case-case genome-wide association analysis shows markers differentially associated with schizophrenia and bipolar disorder and implicates calcium channel genes. Psychiatric Genetics 2011;21:1–4.
Eichler EE, Flint J, Gibson G, Kong A, Leal SM, Moore JH, et al. Missing heritability and strategies for finding the underlying causes of complex disease. Nature Review Genetics 2010;11:446–50.
Fillman SG, Cloonan N, Catts VS, Miller LC, Wong J, McCrossin T, et al. Increased inflammatory markers identified in the dorsolateral prefrontal cortex of individuals with schizophrenia. Molecular Psychiatry 2013;18:206–14.
Gardiner EJ, Cairns MJ, Liu B, Beveridge NJ, Carr V, Kelly B, et al. Gene expression analysis reveals schizophrenia-associated dysregulation of immune pathways in peripheral blood mononuclear cells. Journal of Psychiatric Research 2013;47:425–37.
Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proceedings of the National Academy of Sciences of the United States of the America 2009;106:9362–7.
Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature Protocols 2009;4:44–57.
Ideker T, Ozier O, Schwikowski B, Siegel AF. Discovering regulatory and signalling circuits in molecular interaction networks. Bioinformatics 2002;18(Suppl. 1):S233–240.
Jia P, Wang L, Fanous AH, Pato CN, Edwards TL, Zhao Z. Network-assisted investigation of combined causal signals from genome-wide association studies in schizophrenia. PLOS Computational Biology 2012;8. e1002587.
Jia P, Zheng S, Long J, Zheng W, Zhao Z. dmGWAS: dense module searching for genome-wide association studies in protein-protein interaction networks. Bioinformatics 2011;27:95–102.
Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M. KEGG for representation and analysis of molecular networks involving diseases and drugs. Nucleic Acids Research 2010;38:D355–360.
Kang Y, Zhang X, Dobie F, Wu H, Craig AM. Induction of GABAergic postsynaptic differentiation by alpha-neurexins. Journal of Biological Chemistry 2008;283:2323–34.
Lichtenstein P, Yip BH, Bjork C, Pawitan Y, Cannon TD, Sullivan PF, et al. Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. Lancet 2009;373:234–9.
Liou YJ, Wang HH, Lee MT, Wang SC, Chiang HL, Chen CC, et al. Genome-wide association study of treatment refractory schizophrenia in Han Chinese. PLoS One 2012;7. e33598.
Lu F, Lipka AE, Glaubitz J, Elshire R, Cherney JH, Casler MD, et al. Switchgrass genomic diversity, ploidy, and evolution: novel insights from a network-based SNP discovery protocol. Plos Genetics 2013;9. e1003215.
Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. Nature 2009;461:747–53.
Mouaffak F, Kebir O, Chayet M, Tordjman S, Vacheron MN, Millet B, et al. Association of disrupted in schizophrenia 1 (DISC1) missense variants with ultra-resistant schizophrenia. Pharmacogenomics Journal 2011;11:267–73.
O'Donovan MC, Craddock N, Norton N, Williams H, Peirce T, Moskvina V, et al. Identification of loci associated with schizophrenia by genome-wide association and follow-up. Nature Genetics 2008;40:1053–5.
O'Dushlaine C, Kenny E, Heron E, Donohoe G, Gill M, Morris D, et al. Molecular pathways involved in neuronal cell adhesion and membrane scaffolding contribute to schizophrenia and bipolar disorder susceptibility. Molecular Psychiatry 2011;16:286–92.
Pitcher GM, Kalia LV, Ng D, Goodfellow NM, Yee KT, Lambe EK, et al. Schizophrenia susceptibility pathway neuregulin 1-ErbB4 suppresses Src upregulation of NMDA receptors. Nature Medicine 2011;17:470–8.
Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. American Journal of Human Genetics 2007;81:559–75.
Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature 2009;460:748–52.
Ripke S, Sanders A, Kendler K, Levinson D, Sklar P, Holmans P, et al. Genome-wide association study identifies five new schizophrenia loci. Nature Genetics 2011;43:969–76.
Shi J, Levinson DF, Duan J, Sanders AR, Zheng Y, Pe'er I, et al. Common variants on chromosome 6p22.1 are associated with schizophrenia. Nature 2009;460:753–7.
Stefansson H, Ophoff RA, Steinberg S, Andreassen OA, Cichon S, Rujescu D, et al. Common variants conferring risk of schizophrenia. Nature 2009;460:744–7.
Steinberg S, de Jong S, Andreassen OA, Werge T, Borglum AD, Mors O, et al. Common variants at VRK2 and TCF4 conferring risk of schizophrenia. Human Molecular Genetics 2011;20:4076–81.
Sun YV. Integration of biological networks and pathways with genetic association studies. Human Genetics 2012;131:1677–86.
Thaker GK, Carpenter Jr WT. Advances in schizophrenia. Nature Methods 2001;7:667–71.
Wang K, Li M, Bucan M. Pathway-based approaches for analysis of genomewide association studies. American Journal of Human Genetics 2007;81:1278–83.
Wang K, Li M, Hakonarson H. Analysing biological pathways in genome-wide association studies. Nature Reviews Genetics 2010;11:843–54.
Yue WH, Wang HF, Sun LD, Tang FL, Liu ZH, Zhang HX, et al. Genome-wide association study identifies a susceptibility locus for schizophrenia in Han Chinese at 11p11.2. Nature Genetics 2011;43:1228–31.