Identification of a Predictive Model for Schizophrenia Based on SNPs in a Chinese Population

Zhiying Yang1,2,*, Shun Yao1,2,*, Yichong Xu1, Xiaoking Zhang1,2, Yuan Shi1,2, Lijun Wang1,2, Donghong Cui1,2

1Shanghai Mental Health Center, Shanghai Jiao Tong University School of Medicine, Shanghai, People’s Republic of China; 2Shanghai Key Laboratory of Psychotic Disorders, Shanghai Jiao Tong University School of Medicine, Shanghai, People’s Republic of China

*These authors contributed equally to this work

Correspondence: Donghong Cui, Email manyucc@126.com

Background: Schizophrenia is a devastating mental disease with high heritability. A growing number of susceptibility genes associated with schizophrenia, as well as their corresponding SNPs loci, have been revealed by genome-wide association studies. However, using SNPs as predictors of disease and diagnosis remains difficult. Here, we aimed to uncover susceptibility SNPs in a Chinese population and to construct a prediction model for schizophrenia.

Methods: A total of 210 participants, including 70 patients with schizophrenia, 70 patients with bipolar disorder, and 70 healthy controls, were enrolled in this study. We estimated 14 SNPs using published risk loci of schizophrenia, and used these SNPs to build a model for predicting schizophrenia via comparison of genotype frequencies and regression. We evaluated the efficacy of the diagnostic model in schizophrenia and control patients using ROC curves and then used the 70 patients with bipolar disorder to evaluate the model’s differential diagnostic efficacy.

Results: 5 SNPs were selected to construct the model: rs148415900, rs71428218, rs4666990, rs112222723 and rs1716180. Correlation analysis results suggested that, compared with the risk SNP of 0, the risk SNP of 3 was associated with an increased risk of schizophrenia (OR = 13.00, 95% CI: 2.35–71.84, p = 0.003). The ROC-AUC of this prediction model for schizophrenia was 0.719 (95% CI: 0.634–0.804), with the greatest sensitivity and specificity being 60% and 80%, respectively. The ROC-AUC of the model in distinguishing between schizophrenia and bipolar disorder was 0.591 (95% CI: 0.497–0.686), with the greatest sensitivity and specificity being 60% and 55.7%, respectively.

Conclusion: The SNP risk score prediction model had good performance in predicting schizophrenia. To the best of our knowledge, previous studies have not applied SNP-based models to differentiate between cases of schizophrenia and other mental illnesses. It could have several potential clinical applications, including shaping disease diagnosis, treatment, and outcomes.

Keywords: schizophrenia, SNP, diagnostic model, bipolar disorder, differential diagnosis

Introduction

Schizophrenia is a chronic and complex neuropsychiatric disorder characterized by psychiatric symptoms, behavioral abnormalities, and cognitive impairment. As a severe mental illness, schizophrenia affects about one in a hundred people,1 and is associated with tremendous loss to individuals, families, and society. It is estimated that schizophrenia contributes 13.4 million the years of life lived with disability (YLDs) to the burden of disease globally-equivalent to 1.7% of total global YLDs in 2016.2 Furthermore, the life expectancy of people with schizophrenia is approximately 20 years shorter than average.3

Heredit plays an important role in the onset of mental disorders,4 and the heritability of schizophrenia has been estimated to be as high as 80%.5 Genetic information is expressed through genes, and single nucleotide polymorphisms...
SNPs are base-pair variations at specific positions in the genome. These SNPs can generate variations within risk genes that make the disease more likely. Schizophrenia is the disease with the largest number of within-phenotype SNPs replicated to-date. Additionally, as more genome-wide association studies have been completed in recent years, a growing number of susceptibility genes associated with schizophrenia, as well as their corresponding SNPs, have been found. The largest and most striking among these studies were published by PGC in 2011, 2014, and 2022. They identified more than 300 important SNPs which were mapped to more than 200 loci. These findings provide operability for the precise diagnosis of schizophrenia.

As a disease with polygenic inheritance, schizophrenia is less affected by certain SNPs due to genetic heterogeneity, and prediction models containing several or even dozens of risk SNPs do not always perform well. However, models with more genetic variants included do show better discrimination power for disease prediction. We have improved our understanding of the genetic risks underlying schizophrenia, and paved the way for disease prediction, by establishing polygenic risk score (PRS). However, many SNPs that have been reported to be associated with schizophrenia have not been confirmed across studies, perhaps due to differences in cohort ethnic and/or genetic backgrounds. Thus, using SNPs for diagnostic prediction remains challenging. In particular, no relevant studies have reported on a model that can accurately distinguish between schizophrenia and psychiatric disorders using SNPs.

Bipolar disorder, like schizophrenia, is a severe mental illness. Bipolar and schizophrenic patients often share similar symptoms, treatment responses, and outcomes, and the diseases may even have overlapping pathogeneses and familial patterns. Psychotic disorders have long been defined by clinical features alone, but it is sometimes difficult to clearly distinguish between schizophrenia and bipolar disorder only using symptoms. Thus, more accurate, genetic marker-based score systems for auxiliary and differential diagnosis of schizophrenia are urgently needed.

In this study, we screened several schizophrenia susceptibility SNPs in a Chinese population using published SNPs of schizophrenia risk genes, and determined the optimal combination to construct a risk score model. We then validated our model and applied it to distinguish between schizophrenic and bipolar patients.

Materials and Methods

Participants

Study participants were from the Shanghai Mental Health Center (SMHC) Key Laboratory Precision Medicine Project, and included patients with schizophrenia, patients with bipolar disorder, and healthy individuals. The Project aimed to study severe mental disorders based on genetic, imaging, and clinical features. All patients in our study were recruited from a cohort of psychiatric outpatients and inpatients throughout the country. All participants underwent a comprehensive clinical evaluation, including receiving a validated diagnosis of schizophrenia or bipolar disorder as defined by DSM-IV criteria by two experienced psychiatrists. Healthy controls (HC) were recruited from the community and were also assessed by trained psychiatrists, who confirmed that they had no previous or current mental illnesses. The sample set consisted of participants’ personal information, venous blood, and relevant assessment scales. A synthesized database of genetic, cerebral morphological, neuropsychological, and clinical information was then established. Written informed consent was obtained from all participants. All project procedures were reviewed and approved by the SMHC Ethics Committee.

We selected 70 patients with schizophrenia, 70 patients with bipolar disorder and 70 healthy controls (with a 1:1 ratio of men to women in each group) for a total of 210 participants. Patients who had severe physical diseases or had been diagnosed with any other major Axis I disorder were excluded. Finally, healthy individuals who fulfilled the criteria for any major Axis I disorder or had a family history of mental disorders were excluded.

DNA Extraction

Venous blood was abstracted from all participants, and genomic DNA was extracted with commercially available reagents in accordance with the manufacturer’s protocols (Blood Genomic DNA Extraction Kit, TIANGEN, Beijing). DNA samples were stored at −80 degrees Celsius until they were analyzed.
Selection and Genotyping
The SNPs we selected have statistically significant associations with schizophrenia risk and have been repeatedly validated in GWAS studies or have demonstrated consistency in multiple independent samples. We paid particular attention to those SNPs that show associations in both European and East Asian populations to investigate their possible cross-population effects. All 14 loci were genotyped using the Agena MassARRAY platform (Agena Bioscience, San Diego, CA, USA).

Statistical Analyses
Values conforming to a normal distribution, including age, height, and weight, were expressed as the mean ± standard deviation and independent two-sample t-tests were used to assess differences in continuous variables. Chi-square tests were performed for categorical variables, including marriage, education, and genotypic and allelic frequencies of each SNP. SNPs with a marked deviation from Hardy-Weinberg equilibrium (p<0.05) were excluded from further analysis. Chi-square tests were used to preliminarily explore the differences in genotype and allele frequency of candidate SNPs between schizophrenia patients and healthy controls, and SNPs meeting certain p-value conditions also included in stepwise regression analyses were selected for the candidate SNPs associated with schizophrenia in our Chinese patient population. Odds ratios (ORs) and 95% CIs were calculated to estimate each SNP’s impact on schizophrenia risk. Receiver operating characteristic curves (ROC curves) and area under the curves (AUCs) were used to evaluate the efficacy of risk SNPs in the diagnosis and differential diagnosis of schizophrenia. ROC (Receiver Operating Characteristic) analysis is a technique used to evaluate the performance of a binary classifier. It does so by plotting the true positive rate (also known as sensitivity or recall) against the false positive rate (also known as 1-specificity) at various threshold settings. All tests were performed using SPSS Statistics version 26.0 (IBM), and differences were considered statistically significant at p<0.05 (two-sided).

Results
Demographic Characteristics
Demographic data are shown in Table 1. Although the male-female ratio was fixed, the age distribution was significantly different between patients and controls. The average age was higher in the schizophrenia group than in the HCs (p=0.026). The average weight in the SCZ group was 70.0±13.5 (kg), while the average weight in the HC group was 62.9±10.5 (kg) (p=0.001). Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) total scores were lower in the SCZ cohort than in HCs (p<0.001). There were no significant differences in marriage rates, educational attainment, or height between the two groups.

Table 1 Comparison of Demographic Profiles Between Schizophrenia Patients and Controls

<table>
<thead>
<tr>
<th></th>
<th>SCZ(n=70)</th>
<th>HC(n=70)</th>
<th>t or χ²</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (M ± SD, years)</td>
<td>40.60±12.90</td>
<td>35.90±12.10</td>
<td>2.250</td>
<td>0.026</td>
</tr>
<tr>
<td>Gender (M/F)</td>
<td>35/35</td>
<td>35/35</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Marital status (unmarried/others)</td>
<td>37/33</td>
<td>31/39</td>
<td>1.029</td>
<td>0.310</td>
</tr>
<tr>
<td>Education (=&lt;9 years/&gt;9 years)</td>
<td>16/54</td>
<td>12/58</td>
<td>0.714</td>
<td>0.398</td>
</tr>
<tr>
<td>Height (M ± SD, cm)</td>
<td>166.70±7.80</td>
<td>167.60±8.00</td>
<td>0.677</td>
<td>0.500</td>
</tr>
<tr>
<td>Weight (M ± SD, kg)</td>
<td>70.00±13.50</td>
<td>62.90±10.50</td>
<td>3.437</td>
<td>0.001</td>
</tr>
<tr>
<td>PANSS positive score</td>
<td>14.64±7.69</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>PANSS negative score</td>
<td>14.62±7.56</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>PANSS general psychopathology score</td>
<td>28.93±10.16</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>PANSS total score</td>
<td>58.20±22.12</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>RBANS total score</td>
<td>75.87±17.75</td>
<td>94.69±17.73</td>
<td>6.275</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Abbreviations: SCZ, schizophrenia patients; HC, healthy controls; PANSS, Positive and Negative Syndrome Scale; RBANS, Repeatable Battery for the Assessment of Neuropsychological Status.
Hardy-Weinberg Equilibrium
The genotype frequencies of the 14 candidate SNPs were consistent with Hardy-Weinberg equilibria (both $p>0.05$), indicating that the genotype distribution of the candidate SNPs was balanced, representative of the population, and could be further analyzed (Table 2).

Association Between 14 SNPs and Schizophrenia
Among the 14 selected SNP genotypes, there were significant between-group differences in rs71428218 and rs148415900. The genotype frequencies of rs71428218 in SCZ were CC 78.60%, CT 18.60%, TT 2.90%, but the values were CC 61.40%, CT 37.10%, TT 1.40% for HC ($p=0.044$). For rs148415900, the GG and TG in HC genotype frequencies were 91.40% and 8.60%, respectively, while the frequencies in SCZ were 64.30% and 35.70%, respectively ($p<0.001$) (Supplementary Table 1).

Risk SNPs for Schizophrenia in a Chinese Population
To compare the genotype frequencies of the 14 SNPs between schizophrenia patients and healthy controls, we used chi-square test with different inheritance modes (Supplementary Table 1-Supplementary Table 3). We included SNPs with $p<0.1$, and eventually identified rs148415900, rs71428218, rs4666990, rs112222723, and rs1716180 in the recessive mode as the most sensitive schizophrenia-associated SNPs in the Chinese population (Supplementary Table 3). Further analysis suggested that, compared with a risk SNP of 0, a risk SNP of 3 was associated with an increased risk of schizophrenia (OR=13.00, 95% CI: 2.35–71.84, $p=0.003$). Additionally, as the risk of these SNPs increased, the risk of incidence of schizophrenia also increased (OR=14.40, 95% CI: 2.31–89.94, $p=0.004$). These results survived after adjusting for gender, age, educational attainment, height, and weight (Table 3).

Table 2 Hardy-Weinberg Equilibrium Results

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chromosome:Position</th>
<th>Allele</th>
<th>$X^2$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs2660304</td>
<td>1:98046571</td>
<td>T/G</td>
<td>1.342</td>
<td>0.511</td>
</tr>
<tr>
<td>rs10861879</td>
<td>12:108215857</td>
<td>A/G</td>
<td>1.956</td>
<td>0.376</td>
</tr>
<tr>
<td>rs6804239</td>
<td>3:162062700</td>
<td>T/C</td>
<td>0.071</td>
<td>0.965</td>
</tr>
<tr>
<td>rs9567393</td>
<td>13:32189620</td>
<td>A/G</td>
<td>0.091</td>
<td>0.955</td>
</tr>
<tr>
<td>rs28735056</td>
<td>18:79862879</td>
<td>A/G</td>
<td>0.159</td>
<td>0.924</td>
</tr>
<tr>
<td>rs71428218</td>
<td>13:56374126</td>
<td>C/T</td>
<td>1.176</td>
<td>0.555</td>
</tr>
<tr>
<td>rs148415900</td>
<td>8:64405112</td>
<td>G/T</td>
<td>0.140</td>
<td>0.932</td>
</tr>
<tr>
<td>rs4697446</td>
<td>4:24267999</td>
<td>T/G</td>
<td>0.029</td>
<td>0.986</td>
</tr>
<tr>
<td>rs112222723</td>
<td>10:63102717</td>
<td>T/C</td>
<td>0.391</td>
<td>0.822</td>
</tr>
<tr>
<td>rs7596038</td>
<td>2:58156685</td>
<td>C/T</td>
<td>1.305</td>
<td>0.521</td>
</tr>
<tr>
<td>rs1716180</td>
<td>12:123197534</td>
<td>A/G</td>
<td>0.869</td>
<td>0.647</td>
</tr>
<tr>
<td>rs6983764</td>
<td>8:54828373</td>
<td>G/A</td>
<td>0.032</td>
<td>0.984</td>
</tr>
<tr>
<td>rs4666990</td>
<td>2:184798577</td>
<td>C/T</td>
<td>0.615</td>
<td>0.735</td>
</tr>
</tbody>
</table>

Note: Bolding represents major allele.
Prediction Model Construction
In combination with the risk score results, the five selected SNPs were used as predictors to construct a diagnostic model for schizophrenia, and the diagnostic ability of this prediction model was evaluated. The ROC-AUC of this prediction model was 0.719 (95% CI: 0.634–0.804), with the greatest sensitivity and specificity (of 60% and 80%, respectively) attained at a risk score of 2.5 (Figure 1).

Identification of Bipolar Disorder
70 patients with bipolar disorder were also enrolled, and the prediction model was used for differential diagnosis. The ROC-AUC of the model for distinguishing between schizophrenia and bipolar disorder was 0.591 (95% CI: 0.497–0.686). A risk score of 2.5 was associated with the greatest sensitivity and specificity in this model (60% and 55.7%, respectively) (Figure 2).

Discussion
In this study, 5 SNPs (rs71428218, rs148415900, rs112222723, rs1716180, rs4666990) were selected from 14 risk SNPs that were generated from published SNPs known to be associated with schizophrenia and then used to construct a risk score model for schizophrenia. There was a significant association between schizophrenia and the presence of three or more of the selected SNPs. The ROC-AUC for this polygenic risk score model was 0.719 (95% CI: 0.634–0.804), with a risk score of 2.5 associated with the greatest sensitivity and specificity (60% and 80%, respectively).
more risk SNPs. This risk score model reached high sensitivity and specificity for the diagnosis of schizophrenia and showed good discriminating ability in the diagnosis of bipolar disorder vs schizophrenia.

Previous reports have suggested that expressive suppression is positively associated with psychiatric symptoms in schizophrenia. Our study confirmed a significant positive correlation between the T allele at rs4666990 and blunted effect (Pearson’s \( r=0.276, p=0.021 \)). We also found that rs4666990 was significantly associated with delusions (Pearson’s \( r=0.295, p=0.013 \)), suspiciousness/persecution (Pearson’s \( r=0.255, p=0.033 \)), and lack of judgment and insight (Pearson’s \( r=0.292, p=0.014 \)). rs4666990 is found in ZNF804A (zinc-finger protein 804A), and is one of the strongest susceptibility genes implicated in genome-wide association studies of schizophrenia. It is known to affect protein translation and neural development by encoding multiple proteins. The ZNF804A gene number can regulate multiple aspects of the cellular transcriptome, including cellular adhesion, cell growth, cell size, and leucocyte migration. It also induces alterations in synaptic development and influences neurite outgrowth—a vital component in the pathogenesis of schizophrenia.

Cognitive impairment is the core symptom of schizophrenia. In this study, we explored the association between SNPs and cognition in healthy and diseased individuals. Among patients with schizophrenia, patients with the rs71428218 risk allele(T-) had significantly worse scores on multiple measures of delayed recall (Pearson’s \( r=−0.295, p=0.013 \)), including list recall (Pearson’s \( r=−0.255, p=0.033 \)) and recognition tasks (Pearson’s \( r=−0.243, p=0.042 \)). This difference remained significant even after adjustment for age and educational attainment. This finding is in line with recent studies which have shown that the presence of even one risk SNP can lead to significantly decreased neuronal plasticity.

rs1716180 is part of the MPHOSPH9 gene, which was first associated with neuronal function in 2007, and has more recently been associated with comorbid schizophrenia and type 2 diabetes. rs148415900 and rs112222723 are located on chromosomes 8 and 10, respectively, and although the corresponding genes and functions have not been clearly defined, they have been demonstrated to be susceptibility loci of schizophrenia in European and East Asian genome-wide association studies. Exploring the influence of genes and SNPs on diseases is often achieved through functional pathways in genome-wide association studies, but the function of these SNPs has not been fully explored.

Schizophrenia is a polygenic genetic disorder. After years of genome-wide association studies, many susceptibility SNPs have been found to be associated with schizophrenia, but it can be difficult to accurately distinguish schizophrenia from other mental disorders, such as bipolar disorder, due to overlapping symptoms. Our study was innovative in that
we applied our diagnostic model to both differentiate schizophrenia patients from healthy controls, and to differentiate schizophrenia patients from bipolar patients. To the best of our knowledge, previous studies have not applied SNP-based models to differentiate between cases of schizophrenia and other mental illnesses.

Our study had several limitations. The SNP selection range was limited, and our risk score model should be validated in larger sample sizes. The study primarily focused on the Chinese population, and thus the results may not be applicable to people from other ethnicities or cultural backgrounds. For example, differences in allele frequencies among different ethnic groups may limit the generalizability of the findings. However, our research provides a novel model with the ability to diagnose and distinguish disease-based SNPs, and may eventually have clinical applications. Given that SNP genotypes contain a great deal of information, we imagine that even complex diseases will be able to be accurately diagnosed with SNP genotypes in the future.

**Conclusion**

We found that selecting specific SNPs associated with schizophrenia helped distinguish it from another disease (bipolar disorder) with similar symptoms and pathogenesis. Our SNP-based risk model showed good performance in detecting schizophrenia and in differentiating it from bipolar disorder. This risk score model may have application prospects in clinical medicine.

**Data Sharing Statement**

Data and materials are available and provided within the manuscript or supplementary information files.

**Ethics Approval and Consent to Participate**

All participants signed informed consent. All Project procedures were reviewed and approved by the SMHC Ethics Committee.

**Acknowledgments**

We are grateful to Donghong Cui, Shun Yao for designing the study and proofreading the manuscript. This research was supported by grants from the National Key R&D Program of China (No: 2017YFC0909200), and the National Natural Science Foundation of China (No: 82271544) for Dong Hong Cui; the grant from Shanghai Mental Health Center (2021-YJ09) for Shun Yao. This paper has been uploaded to Research Square as a preprint: https://www.researchsquare.com/article/rs-3919893/v1

**Author Contributions**

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

**Disclosure**

The authors report no conflicts of interest in this work.

**References**


