Loci with genome-wide associations with schizophrenia in the Han Chinese population

Zhiqiang Li,* Yuqian Xiang,* Jianhua Chen, Qiaoli Li, Jiawei Shen, Yun Liu, Wenjin Li, Qinghe Xing, Qingzhong Wang, Lei Wang, Guoyin Feng, Lin He, Xinzhi Zhao and Yongyong Shi

Background
A large schizophrenia genome-wide association study (GWAS) and a subsequent extensive replication study of individuals of European ancestry identified eight new loci with genome-wide significance and suggested that the MIR137-mediated pathway plays a role in the predisposition for schizophrenia.

Aims
To validate the above findings in a Han Chinese population.

Method
We analysed the single nucleotide polymorphisms (SNPs) in the newly identified schizophrenia candidate loci and predicted MIR137 target genes based on our published Han Chinese populations (BIOX) GWAS data. We then analysed 18 SNPs from the candidate regions in an independent cohort that consisted of 3585 patients with schizophrenia and 5496 controls of Han Chinese ancestry.

Results
We replicated the associations of five markers (P < 0.05), including three that were located in the predicted MIR137 target genes. Two loci (ITIH3/4: rs2239547, P = 1.17 × 10⁻¹⁰ and CALN1: rs2944829, P = 9.97 × 10⁻⁶) exhibited genome-wide significance in the Han Chinese population.

Conclusions
The ITIH3/4 locus has been reported to be of genome-wide significance in the European population. The successful replication of this finding in a different ethnic group provides stronger evidence for the association between schizophrenia and ITIH3/4. We detected the first genome-wide significant association of schizophrenia with CALN1, which is a predicted target of MIR137, and thus provide new evidence for the associations between MIR137 targets and schizophrenia.

Declaration of interest
None.

Copyright and usage
© The Royal College of Psychiatrists 2015.

Schizophrenia is a chronic, severe and disabling psychiatric disorder that affects 1% of the population worldwide. Schizophrenia is a complex major disease that manifests as psychotic behaviour (delusions and hallucinations), disorganisation, dysfunction in normal affective responses and altered cognitive functioning. Twin, family and adoption studies have demonstrated that schizophrenia is highly heritable; the heritability of schizophrenia is estimated to be as high as 80%. However, the genetic loci that contribute to the disease remain generally elusive. Recent genome-wide association studies (GWAS) have provided unbiased assessments of common sequence variations across the whole genome and may robustly map the loci involved in the pathology of complex diseases. In the past 5 years, a number of GWAS of schizophrenia have been published. Several loci have surpassed the genome-wide significance threshold (P = 5 × 10⁻⁵) in more than one GWAS; for example the major histocompatibility complex (MHC) region has exhibited genome-wide significance in four independent large-scale GWAS of schizophrenia.

Recently, the Schizophrenia Psychiatric GWAS Consortium (PGC) reported a large schizophrenia GWAS. In this study, the authors conducted a mega-analysis of the combined genotyping data from 21,856 individuals of European ancestry from 17 separate studies and then performed a replication study in a sample of 3585 participants from 19 populations. This study revealed seven loci with genome-wide significance, including two previously reported loci (i.e. the MHC region and TCP4) and the following five novel loci: 1p23.3 (MIR137), 2q32.3 (PCGGEM1), 8p23.2 (CSMD1), 8q21.3 (MMP16), and 10q24.32–q24.33 (CNNM2/NT5C2). Moreover, a joint analysis of schizophrenia and bipolar disorder identified three additional genes that reached genome-wide significance: CACNA1C, ANK3 and ITIH3/4. The most significant new finding was the identification of MIR137, which encodes the microRNA 137, which is a known regulator of neuronal development. Notably, microRNA 137 may directly regulate some other schizophrenia susceptibility genes. Among the 301 high-confidence predicted MIR137 targets, 17 had at least one significant single nucleotide polymorphism (SNP) at P < 10⁻⁵; these 17 targets included four genome-wide significant genes (i.e. TCF4, CACNA1C, CSMD1 and ITIH3/4). Subsequently, these four genes and ZNF804A, another compelling candidate gene for schizophrenia, were validated as MIR137 targets. These findings suggest that the MIR137-mediated pathway is involved in the aetiology of schizophrenia. In a replication study, Hamshere et al tested 78 of the 81 SNPs highlighted by the PGC in a UK population and found significant association for 37 (47%) of the SNPs. Remarkably, genetic variants in three new loci (i.e. ITIH3/4, CACNA1C and SDCCAG8) reached genome-wide significance after combining their new schizophrenia data with those of the PGC. The CLOZUK sample is a series of 2640 UK individuals that were registered for clozapine treatment and had clinical diagnoses of schizophrenia and 2878 controls.

Only a few schizophrenia GWAS have been conducted in non-Western populations. One of these was our GWAS of the genotypes of 3750 individuals with schizophrenia and 6468 healthy controls from Han Chinese populations (BIOX GWAS). None of the SNPs within the 10 loci reported by PGC met the criteria for genome-wide significance in our data-set. However, the criteria for genome-wide significance minimises the occurrence of false positives (i.e. type I errors), whereas the occurrence of false negatives (i.e. type II errors) was almost certain. In another schizophrenia GWAS that was conducted in a Han Chinese population and found significant association for 37 (47%) of the SNPs. Remarkably, genetic variants in three new loci (i.e. ITIH3/4, CACNA1C and SDCCAG8) reached genome-wide significance after combining their new schizophrenia data with those of the PGC. The CLOZUK sample is a series of 2640 UK individuals that were registered for clozapine treatment and had clinical diagnoses of schizophrenia and 2878 controls.

Only a few schizophrenia GWAS have been conducted in non-Western populations. One of these was our GWAS of the genotypes of 3750 individuals with schizophrenia and 6468 healthy controls from Han Chinese populations (BIOX GWAS). None of the SNPs within the 10 loci reported by PGC met the criteria for genome-wide significance in our data-set. However, the criteria for genome-wide significance minimises the occurrence of false positives (i.e. type I errors), whereas the occurrence of false negatives (i.e. type II errors) was almost certain. In another schizophrenia GWAS that was conducted in a Han Chinese population and found significant association for 37 (47%) of the SNPs. Remarkably, genetic variants in three new loci (i.e. ITIH3/4, CACNA1C and SDCCAG8) reached genome-wide significance after combining their new schizophrenia data with those of the PGC. The CLOZUK sample is a series of 2640 UK individuals that were registered for clozapine treatment and had clinical diagnoses of schizophrenia and 2878 controls.
population, Yue et al. identified the susceptibility locus at the MHC region. We ascertained significant associations between the MHC region, the TC4 gene and schizophrenia in our previous study of 2496 patients with schizophrenia and 5184 normal controls drawn from a Han Chinese population, but we failed to confirm an association between the NRG1 gene and schizophrenia. 25 Regarding the psychiatric traits, most of the associations were population specific; however, in some cases, the associations may exhibit convergence of risk genes (but not necessarily risk alleles) across populations. 26 It is of great interest to replicate the findings related to the other loci that were identified by the PGC in the Han Chinese population. In the present study, we conducted a two-stage analysis. We first analysed the SNPs (n = 2595) of the eight newly identified loci, predicted the targets of microRNA 137 in the BIOX GWAS data and selected 18 candidate SNPs. These SNPs were genotyped and tested for their associations with schizophrenia in a replication cohort that consisted of 3585 patients with schizophrenia (the schizophrenia group) and 5496 controls (the control group) of Han Chinese ancestry. A meta-analysis was performed to combine the Chinese data-sets (BIOX GWAS and replication).

Method

Participants

Participants in the schizophrenia group were in-patients or outpatients who were recruited from various mental health centres. The patients were interviewed by two independent psychiatrists and were diagnosed according to DSM-IV criteria, 27 and had 2-year histories of the disorder. All met the following two criteria: preoccupation with one or more delusions and frequent auditory hallucinations. However, none of the following symptoms were prominent: disorganised speech, disorganised or catatonic behaviour, or flat or inappropriate affect. All healthy controls were randomly selected from Chinese Han volunteers (from hospitals and a community survey) who were asked to reply to a written invitation to evaluate their medical histories. Potential lists of controls were screened for suitable volunteers by excluding individuals with major mental illnesses. The sample consisted of 3585 people in the schizophrenia group (1901 men and 1684 women, the mean age at onset of schizophrenia was 35.0 years, s.d. = 11.0) and 5496 controls (2819 men and 2677 women with a mean age of 46.4 years, s.d. = 14.1). Of the participants, 1329 in the schizophrenia group and 2037 in the control group were males, the mean age at onset of schizophrenia was 35.0 years,

BIOX GWAS quality control

The gender established via data genotyping was checked for each of the participants, and individuals in the schizophrenia group of unknown or inconsistent gender (compared with the sample record) were removed (n = 49). Arrays with call rates <95% were excluded (n = 276). SNPs with call rates <95% in either the schizophrenia or the control group were removed (n = 92 324). SNPs with minor allele frequencies <3% (n = 228 267) and those that significantly deviated from Hardy–Weinberg Equilibrium (HWE; P < 1 × 10−4) among the control group (n = 28 657) were also excluded. Heterozygosity rates were calculated with the intent of removing deviations that exceeded six standard deviations from the mean; however, no samples were excluded based on this criterion. PLINK’s identity by descent analysis was used to detect cryptic relatedness. When a pair of individuals exhibited a PI_HAT > 0.25, the member of the pair with the lower call rate was excluded from the analysis (n = 146). Population ancestry assessments were evaluated using principal components analysis, and all of the samples were of Han Chinese ancestry.

SNP selection and genotyping

The allelic frequencies of the nine genome-wide significant SNPs reported in the PGC and PGC+CLOZUK studies differ widely between the European and Chinese populations (see online Table DS1). In the PGC study, the strongest association signal was observed in locus 1p21.3 (MIR137, rs1625579, P = 1.59 × 10−11).
rs1625529 was not genotyped in the BIOX GWAS data-set; however, a proxy SNP (rs1198588, $D’ = 1$ and $r^2 = 1$ in CHB) was present. This proxy SNP exhibited marginal significance ($P = 6.91 \times 10^{-3}$), and the direction of the effect size was consistent with the PGC report. Therefore, the rs1198588 SNP was selected in the follow-up stage. In 10q24.33 (NT5C2), the reported genome-wide significant SNP was rs11191580 ($P = 1.11 \times 10^{-6}$) in the PCG study. In the follow-up phase, we selected rs732998 ($P = 9.15 \times 10^{-3}$ in the BIOX GWAS data-set), which is in complete linkage disequilibrium with rs11191580 ($D’ = 1$, $r^2 = 1$) in the HAPMAP CHB data-set. Therefore, the rs1198588 SNP was selected in the follow-up stage. In the direction of the effect size was consistent with the PGC report.

In the loci with multiple markers that were found to be significant in the BIOX GWAS data-set, all of the SNPs with $P$-values below 0.01 were selected. For each locus, at least one SNP was selected with the exception of locus 2q32.3 because no significant SNPs were observed in this locus. In the loci with multiple markers that were found to be significant in the BIOX GWAS data-set, all of the SNPs with $P$-values below 0.01 were selected. For each locus, at least one SNP was selected with the exception of locus 2q32.3 because no significant SNPs were observed in this locus. In the BIOX GWAS data-set and because, as reported in the PGC study, rs1696826 was non-polymorphic in the HAPMAP CHB population.

### Results for the 17 top-scoring MIR137 predicted target genes in the BIOX GWAS data-set

Among the 17 MIR137 predicted target genes (i.e., C10orf26, TCF4, CSMD1, CACNA1C, SLC12A2, CALN1, GRIA1, ST13, CADPS2, LUZP2, GLIS2, EPHA7, CADPS2, RGS6, TBC1D12, FAM78A and C20orf108), a total of 15 (i.e. all but CSDC2 and GLIS2) with at least one SNP were genotyped in the BIOX GWAS data-set. Within these regions, 2230 genotyped SNPs were analysed and 29 SNPs (1.30%) at CSMD, CACNA1C, CALN1, CADPS2 and RGS6 exhibited significance at the level of $P < 0.01$. Ten tagsSNPs were selected for the follow-up phase.

### Results of the follow-up phase and combined analysis

The detailed SNP quality control information is given in Table D52, and the full results for all SNPs are listed in online Table D53. In the follow-up study, we replicated the associations of five genes in the BIOX GWAS data-set. Within these regions, 2230 genotyped SNPs were analysed and 29 SNPs (1.30%) at CSMD, CACNA1C, CALN1, CADPS2 and RGS6 exhibited significance at the level of $P < 0.01$. Ten tagsSNPs were selected for the follow-up phase.

<table>
<thead>
<tr>
<th>Locus</th>
<th>Gene</th>
<th>Chromosome</th>
<th>SNP</th>
<th>Position</th>
<th>Minor allele</th>
<th>$P$ (CHB)</th>
<th>$P$ (BIOX)</th>
<th>$P$ (meta-analysis)</th>
<th>$P$ (OR for A1 (95% CI))</th>
<th>Replication</th>
<th>Meta-analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>ITIH3/4</td>
<td>17p13.3</td>
<td>rs239547</td>
<td>52855229</td>
<td>C</td>
<td>1.17 x 10^{-6}</td>
<td>0.81 (0.76-0.87)</td>
<td>5.16 x 10^{-10}</td>
<td>1.12 (1.01-1.22)</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>6</td>
<td>CADPS2</td>
<td>10q24.31</td>
<td>rs2944829</td>
<td>71426572</td>
<td>C</td>
<td>0.84 x 10^{-5}</td>
<td>0.86 (0.79-0.94)</td>
<td>9.15 x 10^{-6}</td>
<td>1.10 (1.01-1.15)</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>7</td>
<td>ITIH3/4</td>
<td>17p13.3</td>
<td>rs2192017</td>
<td>122054147</td>
<td>C</td>
<td>3.94 x 10^{-4}</td>
<td>0.86 (0.79-0.94)</td>
<td>3.62 x 10^{-6}</td>
<td>1.09 (1.05-1.15)</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>10</td>
<td>CADPS2</td>
<td>10q24.31</td>
<td>rs10748844</td>
<td>105314100</td>
<td>A</td>
<td>3.17 x 10^{-6}</td>
<td>1.20 x 10^{-5}</td>
<td>9.97 x 10^{-10}</td>
<td>1.10 (1.01-1.15)</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>10</td>
<td>CADPS2</td>
<td>10q24.31</td>
<td>rs11191580</td>
<td>105314100</td>
<td>C</td>
<td>1.21 x 10^{-3}</td>
<td>1.20 x 10^{-5}</td>
<td>9.97 x 10^{-10}</td>
<td>1.10 (1.01-1.15)</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

a. The position is based on the National Center for Biotechnology Information (NCBI) Genome browser build 36.1. rs2239547 was not available in our GWAS sample.

b. A1, minor allele name (based on the whole sample).

c. The SNP quality control information is given in Table D52.
the European-ancestry populations.\textsuperscript{13} The HapMap data-set revealed that rs12699131 was in tight linkage disequilibrium with rs2944829 in the CEU+TSI (Toscani in Italia) populations ($D^* = 0.94$, $r^2 = 0.68$) and that the frequencies of the rs12699131 and rs2944829 haplotypes of the risk allele ($A-G$) and the protective allele ($G-A$) were 38.8% and 52.2% respectively. These results suggested that our finding of an association (rs2944829) is consistent with the findings (rs12699131) from individuals of European descent.

Of the eight genome-wide significant schizophrenia candidate loci identified in the European populations, the associations at \textit{ITIH3/4} (genomic variance signiﬁcance) and \textit{CACA1} [\textit{P} < 0.001] were replicated in our study. At locus 1p21.3, rs1198588 (MIR137) did not exhibit a significant association in the follow-up study. However, variants in three \textit{MIR137} target genes (i.e., \textit{CALN1}, \textit{CDAP2} and \textit{CACA1}) were associated with schizophrenia in the Han Chinese population at \textit{P} < 0.001, and the variant in \textit{CALN1} attained genome-wide significance.

Discussion

Main findings

To validate the ﬁndings of the PGC and PGC+CLOZUK schizophrenia studies (i.e., the European GWAS data-sets) in the Han Chinese population, we selected 18 SNPs from 12 schizophrenia candidate genes based on the BIOX GWAS data and then performed an association study in an extended Han Chinese replication cohort. We replicated the associations of 5/18 markers at nominal thresholds (\textit{P} < 0.05) with \textit{P} < 0.001 in the combined analysis of the BIOX GWAS data and the replication, and two markers (rs2239547 at \textit{ITIH3/4}, \textit{P} = 1.17 x 10^{-10} and rs2944829 at \textit{CALN1}, \textit{P} = 9.97 x 10^{-5}) reached genome-wide significance. Both of our association signals are consistent with the European ﬁndings.

Signiﬁcance of our ﬁndings

Rs2239547 occurred in an extensive linkage disequilibrium block that contained many genes.\textsuperscript{24} However, rs2239547 is an intronic SNP of \textit{ITIH4}, and the expression quantitative trait loci (eQTL) data-set showed that it is signiﬁcantly associated with the expression of \textit{ITIH4} ($P = 3.51 \times 10^{-11}$) in lymphoblastoid cell lines from HapMap JPT (Japanese in Tokyo, Japan) and CHB populations.\textsuperscript{30} Moreover, the \textit{ITIH4} protein has been found to be completely absent in patients who have had acute ischaemic strokes.\textsuperscript{31} Moreover, patients with schizophrenia have a greater likelihood of developing stroke than do controls.\textsuperscript{32,33} Thus, for this locus, \textit{ITIH4} is one of the most compelling functional candidates for further study.

We detected the first genome-wide signiﬁcance of \textit{CALN1} in schizophrenia via joint analysis of the BIOX GWAS data. \textit{CALN1} has two conserved EF-hand-type calcium-binding motifs and is exclusively and highly expressed in the brain, which suggests that this gene has a role in calcium signalling in the central nervous system.\textsuperscript{34} Interestingly, a recent genome-wide association analysis that identiﬁed 13 new risk loci for schizophrenia suggested involvement of neuronal calcium signalling in the aetiology of schizophrenia.\textsuperscript{35} Additionally, several patients with \textit{CALN1} deletions have been found to exhibit intellectual disabilities.\textsuperscript{36} Cognitive impairment is an important clinical feature of schizophrenia. We suggest that further research may clarify whether \textit{CALN1} is associated with neurocognitive phenotypes.

\textit{MIR137} and five of its targets (\textit{TCP4}, \textit{CAGNA1C}, \textit{CSMD1}, \textit{Cl1orf26} and \textit{ZNF80A4}) were reported to be associated with schizophrenia in the GWAS of individuals of European ancestry, which suggests that \textit{MIR137}-mediated dysregulation is an important aetiological mechanism of schizophrenia.\textsuperscript{15} In this study, we failed to directly replicate the association between \textit{MIR137} and schizophrenia in the Han Chinese. However, genetic variants in three predicted \textit{MIR137}-target genes (i.e., \textit{CALN1}, \textit{CDAP2} and \textit{CACA1}) were signiﬁcantly associated with schizophrenia; these variants included rs2944829, which reached genome-wide signiﬁcance. Our ﬁndings support the hypothesis that the \textit{MIR137} pathway is involved in the aetiology of schizophrenia.

Within the eight genome-wide signiﬁcant loci that were reported by the PGC or PGC+CLOZUK studies, the genome-wide signiﬁcance of one locus (\textit{ITIH3/4}) was replicated in our study, which provides evidence that overlapping polygenic variation exists between ethnically divergent populations. However, the other loci did not reach genome-wide signiﬁcance. One possible reason for the non-replication of some genes is that the selected proxy SNPs are in extremely low linkage disequilibrium with the original genome-wide signiﬁcant SNPs. These failures may also partially be explained by genetic heterogeneity across the different populations. However, other factors, such as polygenic heterogeneity, should also be considered. Moreover, the strongest signal (\textit{MIR137} locus) reported in the PGC study also failed to be replicated in the CLOZUK study. All of these phenomena highlight the genetic complexity of schizophrenia, and some susceptibility for schizophrenia is likely to result from population-speciﬁc variants. Meta-analyses and mega-analyses across different populations and larger sample sizes would efﬁciently aid the unravelling of this complexity. In this study, we enlarged the sample size to validate the European ﬁndings.

In summary, we independently conﬁrmed the association of schizophrenia with \textit{ITIH3/4}, which was previously reported to have genome-wide signiﬁcance, and identiﬁed the genome-wide signiﬁcance of \textit{CALN1}, which had not previously attained genome-wide signiﬁcance. Replication across different ethnic groups provides stronger evidence for the associations between schizophrenia and these two loci, and their biological mechanisms will become increasingly important for understanding of the aetiology of schizophrenia.

Zhiqiang Li, PhD, Children’s Hospital, Fudan University, Shanghai and Bio-X Institutes, Key Laboratory for the Genetics of Developmental and Neuropsychiatric Disorders (Ministry of Education), Shanghai Jiao Tong University, Shanghai; Yunliang Xiang, PhD, Children’s Hospital, Fudan University, Shanghai; Jianhua Chen, PhD, Bio-X Institutes, Key Laboratory for the Genetics of Developmental and Neuropsychiatric Disorders (Ministry of Education), Shanghai Jiao Tong University, Shanghai; Qiaoli Li, PhD, Children’s Hospital and Institutes of Biomedical Sciences, Fudan University, Shanghai; Jiaxue Shen, PhD, Bio-X Institutes, Key Laboratory for the Genetics of Developmental and Neuropsychiatric Disorders (Ministry of Education), Shanghai Jiao Tong University, Shanghai; Yun Liu, PhD, Institutes of Biomedical Sciences, Fudan University, Shanghai; Wenjin Li, PhD, Bio-X Institutes, Key Laboratory for the Genetics of Developmental and Neuropsychiatric Disorders (Ministry of Education), Shanghai Jiao Tong University, Shanghai; Qinghe Xing, PhD, Institutes of Biomedical Sciences, Fudan University, Shanghai; Qingzhong Wang, PhD, Bio-X Institutes, Key Laboratory for the Genetics of Developmental and Neuropsychiatric Disorders (Ministry of Education), Shanghai Jiao Tong University, Shanghai; Lei Wang, PhD, Institutes of Biomedical Sciences, Fudan University, Shanghai; Guoyin Feng, Shanghai Institute of Mental Health, Shanghai; Lin He, PhD, Children’s Hospital, Fudan University, Shanghai and Bio-X Institutes, Key Laboratory for the Genetics of Developmental and Neuropsychiatric Disorders (Ministry of Education), Shanghai Jiao Tong University, Shanghai; Xinzhili Zhao, PhD, Children’s Hospital, Fudan University, Shanghai; Yongsong Shi, PhD, Bio-X Institutes, Key Laboratory for the Genetics of Developmental and Neuropsychiatric Disorders (Ministry of Education), Shanghai Jiao Tong University, Shanghai; People’s Republic of China.

Correspondence: Yongsong Shi, Bio-X Institutes, Key Laboratory for the Genetics of Developmental and Neuropsychiatric Disorders (Ministry of Education), Shanghai Jiao Tong University, 1954 Huashan Road, Shanghai 200030, People’s Republic of China. Email: shiyongyong@gmail.com; Xinzhili Zhao, Children’s Hospital, Fudan University, 138 Youyuqiao Road, Shanghai 200032, People’s Republic of China. Email: xzhilesu@fudan.edu.cn.

First received 23 Apr 2014, final revision 7 Nov 2014, accepted 11 Dec 2014.
References

8 Walsh T, McClellan JM, McCarthy SE, Addington AM, Pierce SB, Cooper GM, et al. Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. Sci Signaling 2008; 320: 539.
29 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. Am J Hum Genet 2007; 81: 559–75.