Genome-wide association study identifies a susceptibility locus for schizophrenia in Han Chinese at 11p11.2

Wei-Hua Yue1,2,18, Hai-Feng Wang3,4, Liang-Dan Sun5,6,18, Fu-Lei Tang1,2, Zhong-Hua Liu7, Hong-Xing Zhang8,9, Wen-Qiang Li8,9, Yan-Ling Zhang1,2, Yang Zhang1,2, Cui-Cui Ma10, Bo Du11, Li-Fang Wang1,2, Yun-Qing Ren5,6, Yong-Feng Yang8,9, Xiao-Feng Hu1,2, Yi Wang3,4, Wei Deng12, Li-Wen Tan13, Yun-Long Tan14, Qi Chen15, Guang-Ming Xu16, Gui-Gang Yang14, Xian-bo Zuo5,6, Hao Yan1,2, Yan-Yan Ruan1,2, Tian-Lan Lu1,2, Xue Han13, Xiao-Hong Ma12, Yan Wang1,2, Li-Wei Cai1,2, Chao Jin10, Hong-Yan Zhang1,2, Jun Yan1,2, Wei-Feng Mi1,2, Xin-Yong Yin16, Wen-Bin Ma10, Qi Liu1,2, Lan Kang1,2, Wei Sun1,2, Cheng-Ying Pan1,2, Mei Shuang1,2, Fu-De Yang14, Chuan-Yue Wang15, Jian-Li Yang16, Ke-Qing Li11, Xin Ma15, Ling-Jiang Li13, Xin Yu1,2, Qi-Zhai Li17, Xun Huang7, Lu-Xian Lv8,9, Tao Li12, Guo-Ping Zhao3, Wei Huang3,4, Xue-Jun Zhang5,6 & Dai Zhang1,2

To identify susceptibility loci for schizophrenia, we performed a two-stage genome-wide association study (GWAS) of schizophrenia in the Han Chinese population: GWAS: 746 individuals with schizophrenia and 1,599 healthy controls; validation: 4,027 individuals with schizophrenia and 5,603 healthy controls). We identified two susceptibility loci for schizophrenia at 6p21-p22.1 (rs1233710 in an intron of ZKSCAN4, Pcombined = 4.76 × 10−11, odds ratio (OR) = 0.79; rs1635 in an exon of NAKAP, Pcombined = 6.91 × 10−12, OR = 0.78; rs2142731 in an intron of PGBDI, Pcombined = 5.14 × 10−10, OR = 0.79) and 11p12.2 (rs11038167 near the 5‘UTR of TSPAN18, Pcombined = 1.09 × 10−11, OR = 1.29; rs11038172, Pcombined = 7.21 × 10−10, OR = 1.25; rs835784, Pcombined = 2.73 × 10−11, OR = 1.27). These results add to previous evidence of susceptibility loci for schizophrenia at 6p21-p22.1 in the Han Chinese population. We found that NAKAP and ZKSCAN4 were expressed in postnatal day 0 (P0) mouse brain. These findings may lead to new insights into the pathogenesis of schizophrenia.

Schizophrenia (MIM 181500) is a severe mental disorder with a lifetime prevalence of ~1% and estimated heritability of ~64–80%1,2. Previous candidate gene studies have indicated that NRG1 (encoding neuregulin 1, 8p22-p11), DISC1 (encoding disrupted in schizophrenia 1, 1q42.1) and other genes might confer risk for schizophrenia (SZGene, see URLs)3,4. In recent years, GWAS of schizophrenia have identified several susceptibility loci, including ZNF804A (encoding zinc-finger protein 804A, 2q32.1) and genes within the extended major histocompatibility complex (MHC) region (6p21), mainly in populations of European descent5–8. As a result of the genetic heterogeneity of schizophrenia among populations of different ancestry, the susceptibility loci for schizophrenia were not consistently replicated across various studies. The current study aims to identify the genetic factors underlying schizophrenia in the Han Chinese population by using a large two-stage GWAS.

In the first stage of the study, we conducted a GWAS of 768 individuals with schizophrenia and 1,733 control subjects of Han Chinese descent using Illumina Human610-Quad BeadChips. In total, 620,901 SNPs and copy number variation (CNV) probes were genotyped. After SNP- and sample-based quality control filtering (see Online Methods), a total of 493,203 autosomal SNPs in 746 individuals with schizophrenia and 1,599 control subjects were retained for further analyses (Table 1). Principal-components analysis (PCA) using EIGENSTRAT confirmed that all of our GWAS samples came from individuals of Han Chinese ancestry when compared with 206 HapMap Chinese Han, Beijing, (CHB) samples (Supplementary Fig. 1a–d)9. No obvious population stratification was observed between subjects with schizophrenia and controls (genomic control inflation factor λGC = 1.01, Supplementary Fig. 1e–h).

1Institute of Mental Health, Peking University, Beijing, China. 2Ministry of Health Key Laboratory of Mental Health, Peking University, Beijing, China. 3Ministry of Science and Technology of China Key Laboratory of Mental Health and Disease Genomics, Chinese National Human Genome Center at Shanghai, Shanghai, China. 4Institute of Dermatology and Department of Dermatology, No. 1 Hospital, Anhui Medical University, Hefei, China. 5State Key Laboratory of Dermatology, Ministry of Science and Technology, Hefei, China. 6Key Laboratory of Molecular and Developmental Biology, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China. 7Department of Psychiatry of the Second Affiliated Hospital of Xinxiang Medical University, Xinxiang, China. 8Henan Mental Hospital, Henan Key Laboratory of Biological Psychiatry, Xinxiang, China. 9Jinzhong Kangning Hospital, Jinzhou, China. 10Hebei Mental Health Center, Baoding, China. 11Department of Psychiatry, West China Hospital, Sichuan University, Chengdu, China. 12Institute of Mental Health, Second Xiangya Hospital of Central South University, Changsha, China. 13Beijing Huilongguan Hospital, Beijing, China. 14Beijing Anding Hospital, Capital Medical University, Beijing, China. 15Tianjin Anding Hospital of Tianjin Municipality, Tianjin, China. 16Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China. 17Institute of Dermatology, Ministry of Health Key Laboratory of Mental Health, Peking University, Beijing, China. 18These authors contributed equally to this work. Correspondence should be addressed to D.Z. (daizang@bjmu.edu.cn) or W.H. (huangwei@cgc.sh.cn).
We tested each SNP for association with schizophrenia using the Cochran-Armitage trend test performed with PLINK v1.07 (see URLs)\(^\text{10}\). The quantile-quantile plot of the observed \(P\) values showed a clear deviation from the null distribution at the tail end of the distribution (Supplementary Fig. 2a), suggesting that there was minimal overall inflation of genome-wide statistical results due to population stratification. In the initial GWAS, two SNPs gave \(P\) values smaller than our GWAS significance threshold of 0.05/493,203 or \(-10^7\) (Table 1 and Supplementary Fig. 2a).

Association of polymorphisms at 6p21-p22.1 and 11p11.2 with schizophrenia was observed using the 1000 Genomes Project Japanese in Tokyo (JPT), and CHB reference panels and LocusZoom (see URLs) (Fig. 1a,b)\(^\text{11}\).

To do a fast-track replication analysis, we performed a replication study by genotyping 46 promising SNPs in 38 loci (\(P_{\text{GWAS}} < 1 \times 10^{-5}\), minor allele frequency (MAF) > 5\%) in an independent cohort of Han Chinese individuals (4,027 cases and 5,603 controls). In the combined study, the heterogeneity across the two stages was evaluated using the Cochran’s \(Q\) statistic to determine the heterogeneity statistic (\(I^2\)) and \(P\) value\(^\text{12}\). The Mantel-Haenszel and DerSimonian-Laird methods were used to calculate the fixed- and random-effect models, respectively\(^\text{13}\). Six SNPs annotated in the 6p21-p22 and 11p11.2 loci were validated, providing consistent, independent evidence of association in the replication sample and highly significant association in the combined studies that reached genome-wide significance (6p21-p22.1: rs1233710, \(P_{\text{combined}} = 4.76 \times 10^{-11}\), \(OR = 0.79\); rs1635, \(P_{\text{combined}} = 6.91 \times 10^{-12}\), \(OR = 0.78\); rs2142731, \(P_{\text{combined}} = 5.14 \times 10^{-10}\), \(OR = 0.79\) and 11p11.2: rs11038167, \(P_{\text{combined}} = 1.09 \times 10^{-11}\), \(OR = 1.29\); rs11038172, \(P_{\text{combined}} = 7.21 \times 10^{-10}\), \(OR = 1.25\); rs835784, \(P_{\text{combined}} = 2.73 \times 10^{-13}\), \(OR = 1.27\) (Fig. 1a,b, Table 2).

Table 3\(^\text{6-8}\). We performed conditional logistic analyses by using rs6913660 and rs1635 as markers for the association detected in individuals of European and Chinese descent, respectively (see Online Methods). Conditioning for the association effect of rs6913660 had a minimal impact on the association at rs1635 (\(P_{\text{GWAS}} = 4.15 \times 10^{-6}\), \(OR_{\text{GWAS}} = 0.73\)) compared to \(P_{\text{conditional}} = 1.28 \times 10^{-5}\), \(OR_{\text{conditional}} = 0.73\), although controlling the association at rs1635 greatly reduced the association at rs6913660 (\(P_{\text{GWAS}} = 9.11 \times 10^{-6}\), \(OR_{\text{GWAS}} = 0.59\)) compared to \(P_{\text{conditional}} = 0.043\), \(OR_{\text{conditional}} = 0.66\) (Supplementary Table 4). Furthermore, by examining the data from the HapMap project, we found that the associated SNPs identified in populations of European ancestry and those identified in the Chinese population have very different MAFs in the two populations. Taken together, our results are consistent with the existence of different susceptibility variants within the MHC region in individuals of European and Chinese ancestry. We do, however, recognize the complexity of the linkage-disequilibrium LD pattern within the MHC region and the need for further fine-mapping analysis to confirm this finding.

Expression quantitative trait loci (eQTL) analysis using the Sanger Institute Geneve database (see URLs) suggested that all six of our validated SNPs at chr6p21-p22.1 and chr11p11.2 had cis-eQTL effects on nearby genes in 195 HapMap 2 samples (55 Utah residents of Northern and Western European descent (CEU), 42 CHB, 42 JPT and 56 Yoruba in Ibadan (YRI) samples) (Supplementary Table 5). For example, the rs1233710 SNP showed cis-eQTL effects on ZKSCAN4 (\(P = 0.013\)) and ZNF323 (\(P = 0.001\)) mRNA expression. Moreover, rs12214383 (chr. 6: 28.3 Mb, \(P_{\text{GWAS}} = 0.005\) in the current GWAS), within the LD block of the three validated SNPs on 6p21-p22.1 (\(D’ \geq 0.9\)), was shown to have a cis-eQTL effect on ZNF323 mRNA expression in the brain in previous reports\(^\text{14,15}\).

Figure 1 Regional plots of the two loci associated with schizophrenia at 6p21-p22.1 and 11p11.2. For genotyped SNPs passing quality control measures in the GWAS, \(-\log_{10} P\) values are plotted as a function of genomic position (in the UCSC March 2006 human reference sequence, hg18). \(P_{\text{GWAS}}\) and \(P_{\text{combined}}\) represent the \(P\) values from the GWAS and combined analyses, respectively. The most strongly associated SNP is represented by a purple diamond. All other SNPs are color coded according to the strength of LD (as measured by \(r^2\)) with this index SNP. The recombination rate from the CHB HapMap sample is plotted in light blue. The positions of the six SNPs (rs1233710, rs1635, rs2142731, rs11038167, rs11038172 and rs835784) identified in this study are indicated by arrows. Gene annotations were adapted from the UCSC Genome Browser. Regional plots of (a) 6p21-p22.1 and (b) 11p11.2.

Table 1 Summary information for the 4,773 individuals with schizophrenia and 7,202 control individuals in the GWAS and replication study

<table>
<thead>
<tr>
<th>Analyses</th>
<th>Sample size</th>
<th>Mean age (s.d.)</th>
<th>Male/female</th>
<th>Sample size</th>
<th>Mean age (s.d.)</th>
<th>Male/female</th>
</tr>
</thead>
<tbody>
<tr>
<td>GWAS</td>
<td>746</td>
<td>34.5 (8.7)</td>
<td>396/350</td>
<td>1,599</td>
<td>35.8 (7.8)</td>
<td>846/753</td>
</tr>
<tr>
<td>Replication</td>
<td>4,027</td>
<td>31.8 (8.3)</td>
<td>1,929/2,098</td>
<td>5,603</td>
<td>32.0 (9.4)</td>
<td>2,839/2,764</td>
</tr>
<tr>
<td>Total</td>
<td>4,773</td>
<td>33.1 (6.9)</td>
<td>2,325/2,448</td>
<td>7,207</td>
<td>32.3 (8.6)</td>
<td>3,685/3,517</td>
</tr>
</tbody>
</table>

and Supplementary Table 1). The results for the other 40 SNPs and their associated quality control statistics are reported in Supplementary Table 2.

The three validated SNPs within 6p21-p22.1 were located in the extended MHC region (Supplementary Fig. 3), which has been reported to be associated with schizophrenia by previous studies (Supplementary Table 5).
Table 2  Association evidence at 6p21-22.1 and 11p11.2

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chr.</th>
<th>Adjacent gene</th>
<th>Position</th>
<th>Allele*</th>
<th>MAF</th>
<th>Case</th>
<th>Control</th>
<th>P value</th>
<th>OR</th>
<th>95% CI</th>
<th>MAF</th>
<th>Case</th>
<th>Control</th>
<th>P value</th>
<th>OR</th>
<th>95% CI</th>
<th>Combined analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs1233710</td>
<td>6p21</td>
<td>ZKSCAN4</td>
<td>28323425</td>
<td>T/C</td>
<td>0.2594</td>
<td>0.3258</td>
<td>4.14 × 10^-6</td>
<td>0.73</td>
<td>0.2748</td>
<td>0.321</td>
<td>4.09 × 10^-7</td>
<td>0.80</td>
<td>4.76 × 10^-10</td>
<td>0.79</td>
<td>0.74(</td>
<td>-0.83)</td>
<td></td>
</tr>
<tr>
<td>rs1635</td>
<td>6p22.1</td>
<td>NFKAP</td>
<td>28335583</td>
<td>T/G</td>
<td>0.2634</td>
<td>0.3301</td>
<td>4.15 × 10^-6</td>
<td>0.73</td>
<td>0.275</td>
<td>0.3248</td>
<td>5.53 × 10^-8</td>
<td>0.79</td>
<td>6.91 × 10^-12</td>
<td>0.78</td>
<td>0.73(</td>
<td>-0.82)</td>
<td></td>
</tr>
<tr>
<td>rs2142731</td>
<td>6p22.1</td>
<td>PGBD1</td>
<td>28358892</td>
<td>A/G</td>
<td>0.1857</td>
<td>0.2411</td>
<td>2.19 × 10^-5</td>
<td>0.72</td>
<td>0.184</td>
<td>0.2195</td>
<td>9.15 × 10^-7</td>
<td>0.80</td>
<td>5.14 × 10^-10</td>
<td>0.79</td>
<td>0.74(</td>
<td>-0.84)</td>
<td></td>
</tr>
<tr>
<td>rs11038167</td>
<td>11p11.2</td>
<td>TSPAN18</td>
<td>44799710</td>
<td>A/C</td>
<td>0.4812</td>
<td>0.399</td>
<td>1.12 × 10^-7</td>
<td>1.40</td>
<td>0.4534</td>
<td>0.4006</td>
<td>3.28 × 10^-4</td>
<td>1.27</td>
<td>1.09 × 10^-11</td>
<td>1.29</td>
<td>(1.23(</td>
<td>-1.36)</td>
<td></td>
</tr>
<tr>
<td>rs11038172</td>
<td>11p11.2</td>
<td>TSPAN18</td>
<td>44812173</td>
<td>A/G</td>
<td>0.4973</td>
<td>0.424</td>
<td>2.57 × 10^-6</td>
<td>1.34</td>
<td>0.47</td>
<td>0.4187</td>
<td>1.11 × 10^-5</td>
<td>1.23</td>
<td>7.21 × 10^-10</td>
<td>1.25</td>
<td>(1.19(</td>
<td>-1.32)</td>
<td></td>
</tr>
<tr>
<td>rs8835784</td>
<td>11p11.2</td>
<td>TSPAN18</td>
<td>44820394</td>
<td>A/G</td>
<td>0.3432</td>
<td>0.277</td>
<td>4.02 × 10^-6</td>
<td>1.36</td>
<td>0.314</td>
<td>0.268</td>
<td>2.37 × 10^-5</td>
<td>1.25</td>
<td>2.73 × 10^-11</td>
<td>1.27</td>
<td>(1.20(</td>
<td>-1.34)</td>
<td></td>
</tr>
</tbody>
</table>

Chr., chromosome; MAF, minor allele frequency; OR, odds ratio; 95% CI, 95% confidence interval.

*Minor allele/major allele.

At 6p21-p22.1, three genes were identified, including NFKAP (encoding piggyBac transposable element-like), ZKSCAN4 (encoding zinc finger with KRAB and SCAN domains) and PGBD1 (encoding piggyBac transposable element derived 1), which have not been reported to have known functions (Fig. 1a). In humans, the sequence of NFKAP is about 55% homologous to that of NFKAP. The NKAP protein is a transcriptional repressor of tumor necrosis factor (TNF)- and interleukin-1–induced nuclear factor-kB activation. It is associated with the histone deacetylase 3 and Notch corepressor complex, which is required for T cell development.8,17. The validated rs1635 SNP is located in exon 1 of the NFKAP gene, representing a nonsynonymous SNP that results in a T152N substitution in the encoded protein.

Using in situ hybridization methods, we additionally determined that NFKAP mRNA was highly expressed in P0 imprinting control region (ICR) mice in the cortex, hippocampus, ventral lateral nucleus, locus ceruleus and other brain areas (Supplementary Fig. 4a). RNA interference (RNAi)-mediated knockdown of NFKAP expression showed that the NKAP might have a role in the regulation of neuronal migration during early neurodevelopment (Supplementary Fig. 5 and Supplementary Note). We evaluated the effects of the knockdown of CG6066 (NFKAP) in Drosophila melanogaster (51.9% homologous to human NFKAP). Compared to wild-type Drosophila, Drosophila with RNAi-mediated knockdown of CG6066 had defects in morphology (misshapen wings, rough eyes and crooked and twisted metathoracic legs) and synaptic defects at the neuromuscular junction, which included enlarged volume and decreased numbers of synapses (Supplementary Fig. 6). Taken together, these findings indicate that NFKAP might have a role in neurodevelopmental processes.

Another validated SNP, rs1233710, was found in an intron of ZKSCAN4. A sequence encoding the ZKSCAN4 protein, also known as zinc finger protein 307 (ZNF307), was cloned from a human embryonic heart cDNA library. The ZKSCAN4 protein contains a leucine-rich repeat (LRR) domain, a Kruppel-associated box (KRAB) domain and seven C2H2 zinc finger motifs. This protein mainly localizes to the nucleus and is highly expressed in brain. It seems to be a transcriptional repressor that inhibits p53 and p21 transcriptional activity by activating MDM2 and EP300 expression.18. Using in situ hybridization, we found that ZKSCAN4 mRNA was highly expressed in the cortex, paraventriculur nucleus and amygdala of P0 mice and was highly expressed in the hippocampus, cornus ammonis (CA) 1–3 regions and dentate gyrus of P14 mice (Supplementary Fig. 4b). These expression pattern data also suggested that ZKSCAN4 might have a role in the process of high brain function by influencing the postnatal development process.

PGBD1 belongs to the piggyBac transposable element-derived subfamily of proteins of unknown function. It is specifically expressed in the mouse brain, especially in the cortex, olfactory bulb, hippocampus, CA1-CA3 and dentate gyrus (see URLs). PGBD1 has been reported to be a susceptibility gene for schizophrenia and Alzheimer’s disease in individuals of European descent.19-21. In the present study, we also found a validated SNP rs2142731 that is highly associated with schizophrenia in an intron of PGBD1. This finding suggested that PGBD1 might be a biologically important candidate gene for the development of schizophrenia. These findings are in line with previous evidence showing that the immune system and neurodevelopment may have important roles in the pathogenesis of schizophrenia.8,10,21.

TSPAN18 (tetraspanin 18) at 11p11.2, for which the function of the protein is unknown according to currently available information, encodes a member of a large family of tetraspanins that are involved in diverse cellular processes.22 Other tetraspanins are known to form tetraspanin-enriched microdomains and may be involved in the clustering of receptors or cell signaling molecules.23 A previous study found that TSPAN8, in the same family as TSPAN18, was associated with bipolar I disorder.24 However, the potential role of the TSPAN18 protein in the pathogenesis of schizophrenia requires further exploration.

In summary, we performed a GWAS of schizophrenia in the Han Chinese population and identified two new susceptibility loci at 6p21-p22.1 and 11p11.2. Our study not only adds to the known genetic factors that predispose individuals to schizophrenia but also highlights the importance of genetic factors in the disease that should advance understanding of the pathogenesis of schizophrenia.


METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank J. Liu from the Human Genetics, Genome Institute of Singapore for his suggestions for revision of the manuscript. We acknowledge with appreciation the contributions made this work possible. This work was supported by research grants from the National High-Tech Research and Development Program of China (2009AA02Z202), the National Natural Science Foundation of China (30502090, 3087899, 81070187 and 81070188), the National Basic Research Program of China (2007CB512301, 2011CB707805) and the International Science & Technology Cooperation Program of China (2010DFB30820).
Identification of loci associated with schizophrenia by
Common polygenic variation
Schizophrenia susceptibility alleles are enriched for alleles
Common variants on chromosome 6p22.1 are associated with

The authors declare no competing financial interests.

Published online at http://www.nature.com/naturegenetics/.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

Statistical analysis of the GWAS. We carried out the Cochran-Armitage trend test to assess genotype-phenotype association using PLINK 1.07 software. We used quantile-quantile plots to evaluate the overall significance of the genome-wide association results. Using stringent Bonferroni correction, we set our genome-wide significance threshold at $0.05/493,203 = 1.01 \times 10^{-7}$. Statistical analysis of the replication study. Replication analysis in the second stage was done by analyzing the follow-up samples separately and then analyzing the combined sample of all the cases and controls in the two stages. Heterogeneity tests ($I^2$ and $P$ values from $Q$ statistics) between different samples were performed using the previously described method. In general, an $I^2$ of <25% was considered to signify no heterogeneity. $P$ of 30–50% signified moderate heterogeneity and $I^2$ >50% indicated strong heterogeneity. In our analysis, we set the threshold of $I^2$ to be 50% for heterogeneity tests. If $I^2$ was less than 50% ($P > 0.05$), the fixed-effect (Mantel-Haenszel) model was used to combine the results from the two different cohorts; otherwise, the random-effect (DerSimonian-Laird) model was used. All $P$ values from the validation analysis are reported without correction for multiple testing.