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

Wei-Hua Yue[1,2,18], Hai-Feng Wang[3,4], Liang-Dan Sun[5,6,18], Fu-Lei Tang[1,2], Zhong-Hua Liu[7], Hong-Xing Zhang[8,9], Wen-Qiang Li[8,9], Yan-Ling Zhang[1,2], Yang Zhang[1,2], Cui-Cui Ma[10], Bo Du[11], Li-Fang Wang[1,2], Yun-Qing Ren[5,6], Yong-Feng Yang[8,9], Xiao-Feng Hu[1,2], Yi Wang[3,4], Wei Deng[12], Li-Wen Tan[13], Yun-Long Tan[14], Qi Chen[15], Guang-Ming Xu[16], Gui-Gang Yang[14], Xian-bo Zuo[5,6], Hao Yan[1,2], Yan-Yan Ruan[1,2], Tian-Lan Lu[1,2], Xue Han[13], Xiao-Hong Ma[12], Yan Wang[1,2], Li-Wei Cai[1,2], Chao Jin[10], Hong-Yan Zhang[1,2], Jun Yan[1,2], Wei-Feng Mi[1,2], Xian-Yong Yin[5,6], Wen-Bin Ma[10], Qi Liu[1,2], Lan Kang[1,2], Wei Sun[1,2], Cheng-Ying Pan[1,2], Mei Shuang[1,2], Fu-De Yang[14], Chuan-Yue Wang[15], Jian-Li Yang[16], Ke-Qing Li[11], Xin Ma[15], Ling-Jiang Li[13], Xin Yu[1,2], Qi-Zhai Li[17], Xun Huang[7], Lu-Xian Lv[8,9], Tao Li[12], Guo-Ping Zhao[3], Wei Huang[3,4], Xue-Jun Zhang[5,6] & Dai Zhang[1,2]

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*, $P_{combined} = 4.76 \times 10^{-11}$, odds ratio (OR) = 0.79; rs1635 in an exon of *NKAPL*, $P_{combined} = 6.91 \times 10^{-12}$, OR = 0.78; rs2142731 in an intron of *PGBD1*, $P_{combined} = 5.14 \times 10^{-10}$, OR = 0.79) and 11p11.2 (rs11038167 near the 5′ UTR of *TSPAN18*, $P_{combined} = 1.09 \times 10^{-11}$, OR = 1.29; rs11038172, $P_{combined} = 7.21 \times 10^{-10}$, OR = 1.25; rs835784, $P_{combined} = 2.73 \times 10^{-11}$, 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 *NKAPL* 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 descent[5–8]. 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 $\lambda_{GC} = 1.01$, **Supplementary Fig. 1e–h**).

[1]Institute of Mental Health, Peking University, Beijing, China. [2]Ministry of Health Key Laboratory of Mental Health, Peking University, Beijing, China. [3]Shanghai–Ministry of Science and Technology of China Key Laboratory of Health and Disease Genomics, Chinese National Human Genome Center at Shanghai, Shanghai, China. [4]State Key Laboratory of Medical Genomics, Ruijin Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, China. [5]Institute of Dermatology and Department of Dermatology, No. 1 Hospital, Anhui Medical University, Hefei, China. [6]State Key Laboratory of Dermatology, Ministry of Science and Technology, Hefei, China. [7]Key Laboratory of Molecular and Developmental Biology, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China. [8]Department of Psychiatry of the Second Affiliated Hospital of Xinxiang Medical University, Xinxiang, China. [9]Henan Mental Hospital, Henan Key Laboratory of Biological Psychiatry, Xinxiang, China. [10]Jinzhou Kangning Hospital, Jinzhou, China. [11]Hebei Mental Health Center, Baoding, China. [12]Department of Psychiatry, West China Hospital, Sichuan University, Chengdu, China. [13]Institute of Mental Health, Second Xiangya Hospital of Central South University, Changsha, China. [14]Beijing Huilongguan Hospital, Beijing, China. [15]Beijing Anding Hospital, Capital Medical University, Beijing, China. [16]Tianjin Anding Hospital of Tianjin Municipality, Tianjin, China. [17]Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China. [18]These authors contributed equally to this work. Correspondence should be addressed to D.Z. (daizang@bjmu.edu.cn) or W.H. (huangwei@chgc.sh.cn).

Received 7 March; accepted 21 September; published online 30 October 2011; doi:10.1038/ng.979

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

	Cases			Controls		
Analyses	Sample size	Mean age (s.d.)	Male/female	Sample size	Mean age (s.d.)	Male/female
GWAS	746	34.5 (8.7)	396/350	1,599	35.8 (7.8)	846/753
Replication study	4,027	31.8 (8.3)	1,929/2,098	5,603	32.0 (9.4)	2,839/2,764
Total	4,773	33.1 (5.9)	2,325/2,448	7,207	32.3 (8.6)	3,685/3,517

We tested each SNP for association with schizophrenia using the Cochran-Armitage trend test performed with PLINK v1.07 (see URLs)[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. 2**), 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 ~1.01×10^{-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**)[11].

To do a fast-track replication analysis, we performed a replication study by genotyping 46 promising SNPs in 38 loci ($P_{GWAS} < 1 \times 10^{-5}$, 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[12]. The Mantel-Haenszel and DerSimonian-Laird methods were used to calculate the fixed- and random-effect models, respectively[13]. Six SNPs annotated in the 6p21-p22 and 11p11.2 loci were validated, providing independent, consistent evidence of association in the replication sample and highly significant association in the combined studies that reached genome-wide significance (6p21-p22.1: rs1233710, $P_{combined} = 4.76 \times 10^{-11}$, OR = 0.79; rs1635, $P_{combined} = 6.91 \times 10^{-12}$, OR = 0.78; rs2142731, $P_{combined} = 5.14 \times 10^{-10}$, OR = 0.79 and 11p11.2: rs11038167, $P_{combined} = 1.09 \times 10^{-11}$, OR = 1.29; rs11038172, $P_{combined} = 7.21 \times 10^{-10}$, OR = 1.25; rs835784, $P_{combined} = 2.73 \times 10^{-11}$, OR = 1.27 (**Fig. 1a**,**b**, **Table 2** 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 3**)[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 ($P_{GWAS} = 4.15 \times 10^{-6}$, $OR_{GWAS} = 0.73$ compared to $P_{conditional} = 1.28 \times 10^{-5}$, $OR_{conditional} = 0.73$), although controlling the association at rs1635 greatly reduced the association at rs6913660 ($P_{GWAS} = 9.11 \times 10^{-3}$, $OR_{GWAS} = 0.59$ compared to $P_{conditional} = 0.043$, $OR_{conditional} = 0.66$) (**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 Genevar 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, $P_{GWAS} = 0.005$ in the current GWAS), within the LD block of the three validated SNPs on 6p21-p22.1 ($D' > 0.9$), was shown to have a *cis*-eQTL effect on *ZNF323* mRNA expression in the brain in previous reports[14,15].

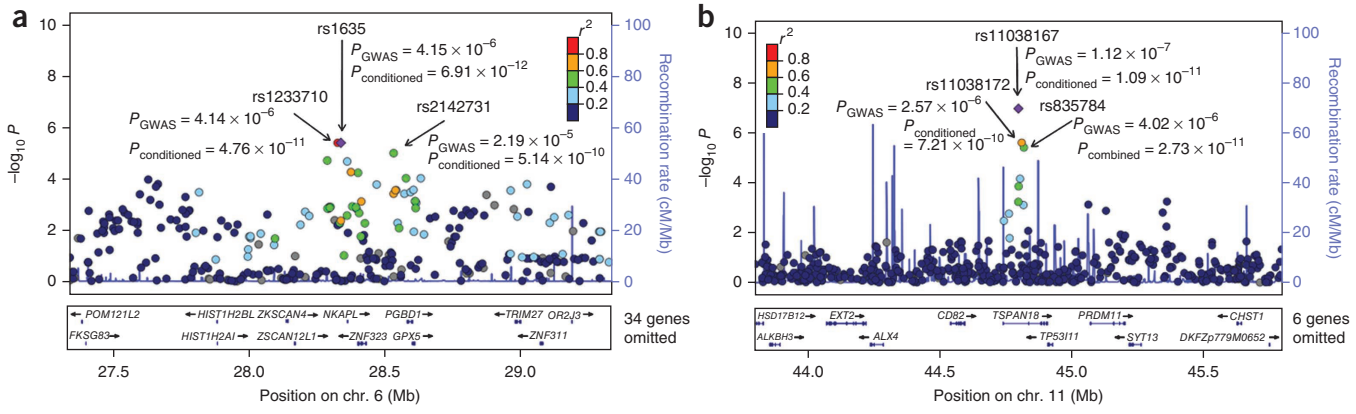

Figure 1 Regional plots of the two loci associated with schizophrenia at 6p21-p22.1 and 11p11.2. For genotyped SNPs passing quality control measures in the GWAS, $-\log_{10} P$ values are plotted as a function of genomic position (in the UCSC March 2006 human reference sequence, hg18). P_{GWAS} and $P_{combined}$ 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 2 Association evidence at 6p21-22.1 and 11p11.2

					GWAS (746 cases, 1,599 controls)				Replication study (4,027 cases, 5,603 controls)				Combined analysis	
					MAF				MAF					
SNP	Chr.	Adjacent gene	Position	Allele[a]	Case	Control	*P* value	OR	Case	Control	*P* value	OR	$P_{combined}$	OR (95% CI)
rs1233710	6p21	*ZKSCAN4*	28323425	T/C	0.2594	0.3258	4.14×10^{-6}	0.73	0.2748	0.321	4.09×10^{-7}	0.80	4.76×10^{-11}	0.79 (0.74–0.83)
rs1635	6p22.1	*NKAPL*	28335583	T/G	0.2634	0.3301	4.15×10^{-6}	0.73	0.275	0.3248	5.53×10^{-8}	0.79	6.91×10^{-12}	0.78 (0.73–0.82)
rs2142731	6p22.1	*PGBD1*	28358892	A/G	0.1857	0.2411	2.19×10^{-5}	0.72	0.184	0.2195	9.15×10^{-7}	0.80	5.14×10^{-10}	0.79 (0.74–0.84)
rs11038167	11p11.2	*TSPAN18*	44799710	A/C	0.4812	0.399	1.12×10^{-7}	1.40	0.4534	0.4006	3.28×10^{-6}	1.27	1.09×10^{-11}	1.29 (1.23–1.36)
rs11038172	11p11.2	*TSPAN18*	44812173	A/G	0.4973	0.424	2.57×10^{-6}	1.34	0.47	0.4187	1.11×10^{-5}	1.23	7.21×10^{-10}	1.25 (1.19–1.32)
rs835784	11p11.2	*TSPAN18*	44820394	A/G	0.3432	0.277	4.02×10^{-6}	1.36	0.314	0.268	2.37×10^{-7}	1.25	2.73×10^{-11}	1.27 (1.20–1.34)

Chr., chromosome; MAF, minor allele frequency; OR, odds ratio; 95% C.I., 95% confidence interval.
[a]Minor allele/major allele.

At 6p21-p22.1, three genes were identified, including *NKAPL* (encoding NFKB activating protein-like), *ZKSCAN4* (encoding zinc finger with KRAB and SCAN domains 4) and *PGBD1* (encoding piggyBac transposable element derived 1), which have not been reported to have known functions (**Fig. 1a**). In humans, the sequence of *NKAPL* is about 55% homologous to that of *NKAP*. The NKAP protein is a transcriptional repressor of tumor necrosis factor (TNF)- and interleukin-1–induced nuclear factor-κB activation. It is associated with the histone deacetylase 3 and Notch corepressor complex, which is required for T cell development[16,17]. The validated rs1635 SNP is located in exon 1 of the *NKAPL* gene, representing a nonsynonymous SNP that results in a T152N substitution in the encoded protein.

Using *in situ* hybridization methods, we additionally determined that *NKAPL* mRNA was highly expressed in P0 imprinting control region (ICR) mice in the cortex, hippocampus, ventral lateral nucleus, locus cerulus and other brain areas (**Supplementary Fig. 4a**). RNA interference (RNAi)-mediated knockdown of *NKAPL* expression showed that the *NKAPL* 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 (NKAPL) in *Drosophila melanogaster* (51.9% homologous to human *NKAPL*). 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 *NKAPL* 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 Kreuppel-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, paraventricular 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[7,19]. 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[20,21].

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.

URLs. SZGene, http://www.szgene.org/; UCSC Genome Browser, http://genome.ucsc.edu/; PLINK, http://pngu.mgh.harvard.edu/~purcell/plink/; LocusZoom, http://csg.sph.umich.edu/locuszoom/; Sanger Institute Genevar database, ftp://ftp.sanger.ac.uk/pub/genevar/; Allen Brain Atlas, http://mouse.brain-map.org/welcome.do.

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 all the individuals with schizophrenia and healthy control subjects whose contributions made this work possible. This work was supported by research grants from the National High-Tech Research and Development Program of China (2009AA022702), the National Natural Science Foundation of China (30530290, 30870896, 81071087 and 81071088), the National Basic Research Program of China (2007CB512301, 2011CB707805) and the International Science & Technology Cooperation Program of China (2010DFB30820).

AUTHOR CONTRIBUTIONS
D.Z., W.H., X.-J.Z., G.-P.Z. and T.L. designed the study. D.Z. and X.-J.Z. revised the manuscript. D.Z. and W.-H.Y. obtained financial support. W.-H.Y., L.-D.S. and L.-F.W. prepared the manuscript. H.-F.W., W.-H.Y. and L.-D.S. supervised the experiments and data analysis. H.-X.Z., W.-Q.L., Y.-L.Z., C.-C.M., B.D., Y.-Q.R., Y.-F.Y., X.-F.H., Y.W., W.D., L.-W.T., Y.-L.T., Q.C., G.-M.X., G.-G.Y., H.Y., Y.-Y.R., T.-L.L., X.H., X.-H.M., Y.W., L.-W.C., C.J., H.-Y.Z., J.Y., W.-F.M., X.-Y.Y., W.-B.M., Q.L., L.K., W.S., C.-Y.P., M.S., F.-D.Y., C.-Y.W., J.-L.Y., K.-Q.L., X.M., L.-J.L., X.Y. and L.-X.L. conducted sample selection and data management, undertook recruitment, collected phenotype data, undertook related data handling and calculation, managed recruitment and obtained biological samples. W.-H.Y., L.-F.W., X.B.Z. and Q.-Z.L. undertook data processing, statistical analysis and bioinformatics investigations. F.-L.T., Z.-H.L., Y.Z. and X.H. performed *in situ* hybridization and RNAi experiments. All authors critically reviewed the manuscript and approved the final version.

COMPETING FINANCIAL INTERESTS
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.

1. Lichtenstein, P. *et al.* Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. *Lancet* **373**, 234–239 (2009).
2. Thaker, G.K. & Carpenter, W. Advances in schizophrenia. *Nat. Med.* **7**, 667–671 (2001).
3. Harrison, P.J. & Owen, M.J. Genes for schizophrenia? Recent findings and their pathophysiological implications. *Lancet* **361**, 417–419 (2003).
4. Allen, N.C. *et al.* Systematic meta-analyses and field synopsis of genetic association studies in schizophrenia: the SzGene database. *Nat. Genet.* **40**, 827–834 (2008).
5. O'Donovan, M.C. *et al.* Identification of loci associated with schizophrenia by genome-wide association and follow-up. *Nat. Genet.* **40**, 1053–1055 (2008).
6. International Schizophrenia Consortium *et al.* Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748–752 (2009).
7. Stefansson, H. *et al.* Common variants conferring risk of schizophrenia. *Nature* **460**, 744–747 (2009).
8. Shi, J. *et al.* Common variants on chromosome 6p22.1 are associated with schizophrenia. *Nature* **460**, 753–757 (2009).
9. Price, A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
10. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
11. Pruim, R.J. *et al.* LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26**, 2336–2337 (2010).
12. Higgins, J.P. & Thompson, S.G. Quantifying heterogeneity in a meta-analysis. *Stat. Med.* **21**, 1539–1558 (2002).
13. Mantel, N. & Haenszel, W. Statistical aspects of the analysis of data from retrospective studies. *J. Natl. Cancer Inst.* **22**, 719–748 (1959).
14. Gibbs, J.R. *et al.* Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. *PLoS Genet.* **6**, e1000952 (2010).
15. Richards, A.L. *et al.* Schizophrenia susceptibility alleles are enriched for alleles that affect gene expression in adult human brain. *Mol. Psychiatry* (in the press).
16. Pajerowski, A.G. *et al.* NKAP is a transcriptional repressor of Notch signaling and is required for T cell development. *Immunity* **30**, 696–707 (2009).
17. Chen, D. *et al.* Identification of a nuclear protein that promotes NF-κB activation. *Biochem. Biophys. Res. Commun.* **310**, 720–724 (2003).
18. Li, J. *et al. ZNF307*, a novel zinc finger gene suppresses p53 and p21 pathway. *Biochem. Biophys. Res. Commun.* **363**, 895–900 (2007).
19. Bertram, L. & Tanzi, R.E. Genome-wide association studies in Alzheimer's disease. *Hum. Mol. Genet.* **18**, R137–R145 (2009).
20. Strous, R.D. & Shoenfeld, Y. Schizophrenia, autoimmunity and immune system dysregulation: a comprehensive model updated and revisited. *J. Autoimmun.* **27**, 71–80 (2006).
21. McGlashan, T.H. & Hoffman, R.E. Schizophrenia as a disorder of developmentally reduced synaptic connectivity. *Arch. Gen. Psychiatry* **57**, 637–648 (2000).
22. Hemler, M.E. Tetraspanin proteins mediate cellular penetration, invasion, and fusion events and define a novel type of membrane microdomain. *Annu. Rev. Cell Dev. Biol.* **19**, 397–422 (2003).
23. Berditchevski, F. & Odintsova, E. Tetraspanins as regulators of protein trafficking. *Traffic* **8**, 89–96 (2007).
24. Sklar, P. *et al.* Whole-genome association study of bipolar disorder. *Mol. Psychiatry* **13**, 558–569 (2008).

ONLINE METHODS

Subjects. In this study, we performed two-stage case-control association analyses in two samples that were independent of each other. The initial GWAS sample consisted of 768 unrelated subjects with schizophrenia (408 males and 360 females) and 1,733 control subjects (917 males and 816 females). For validation, an independent sample consisting of 4,027 cases (1,929 males and 2,098 females) and 5,603 controls (2,839 males and 2,764 females) was recruited from northern China. All the cases and controls were determined to be of northern Han Chinese origin because they and their parents were born in Beijing, Tianjin, Hebei, Shandong, Henan or Anhui provinces. The genetic structure of the Chinese population indicated that there is a subpopulation structure between the southern and northern Han Chinese populations and that geographic location could be a good indicator of ancestral origin for genetic matching[25,26]. Consensus diagnoses were made by at least two experienced psychiatrists according to the criteria for schizophrenia from the Diagnostic and Statistical Manual of Mental Disorders IV (DSM-IV). No subjects had severe medical complications or other psychiatric disorders. The control samples for the GWAS and related data were supplied by the State Key Laboratory of Dermatology, Ministry of Science and Technology in Hefei, China. All control individuals were clinically determined to be free of autoimmune or psychiatric disorders or family history of such disorders (including first-, second- and third-degree relatives). Both case and control groups were matched according to age and gender (**Table 1**). This study was approved by the Institutional Review Board of each hospital. Written informed consent was obtained from all participants.

Sample preparation and genotyping. Genomic DNA samples were extracted using the QIAamp DNA Mini Kit (QIAGEN). The genotyping of denatured samples in the first stage of the GWAS was performed on Illumina HumanHap610-Quad BeadChips, which include 620,901 SNPs and CNV probes in total. After hybridization, the BeadChip oligonucleotides were extended by a single labeled base, which was detected by fluorescence imaging with an Illumina Bead Array Reader. Normalized bead intensity data obtained from samples were loaded into Illumina BeadStudio 3.2 software, which converted fluorescence intensities into SNP genotypes. On the basis of the initial GWAS results, 46 SNPs were selected for fast-track validation analysis, and their genotypes were determined using the Sequenom MassARRAY system (Sequenom iPLEX, primer sequences provided in **Supplementary Table 6**). Genotyping was performed at the State Key Laboratory of Dermatology, Ministry of Science and Technology in Hefei, China according to the manufacturer's instructions. An additional validation panel was genotyped using a TaqMan genotyping platform (ABI 7300 Real Time PCR system, Applied Biosystems) for rs1801133 and rs6978425. The concordance between DNA sequencing (six SNPs in 100 randomly selected samples) on an ABI PRISM 377-96 DNA Sequencer (Applied Biosystems) and genotyping performed on Illumina BeadChips and by Sequenom MassARRAY assays was >99%.

Statistical analysis of the GWAS and replication studies. *Quality control.* Initially, 768 cases and 1,733 controls were genotyped with 620,901 SNPs and CNV probes. As a part of quality control, we also examined potential genetic association based on pairwise identity-by-state analysis for all the successfully genotyped samples. Upon identification of a pair containing first- or second-degree relatives, we removed one of the two related individuals (the sample with the lower call rate). One case and two controls were removed for poor genotyping (MIND >0.1) or relative relationship. After stringent quality control assessment, we excluded 30,164 SNPs with call rates <90%, 111,392 SNPs with minor allele frequencies <1% and 3,265 SNPs with significant deviation from Hardy-Weinberg equilibrium ($P < 1 \times 10^{-5}$) in the controls. For all of the 620,901 SNPs and CNV probes in 768 cases and 1,733 healthy control subjects, the total genotyping rate was 0.952601. After quality control filtering, the total genotyping rate was 0.999291 in the remaining individuals (746 cases and 1,599 controls) and 493,203 remaining markers. Of our top 40 associated SNPs, none showed significant differences in missing rates between cases and controls in the GWAS ($P > 0.05$).

Principal-components analysis. We performed PCA following the methodology of the EIGENSTRAT software package in two steps (**Supplementary Fig. 1**)[9]. First, PCA was used to identify genetic outliers. We combined 2,498 participant samples (after removal of samples with low call rates or familial relationships) with 206 HapMap samples and performed PCA. The HapMap samples were drawn from the 57 YRI, 44 JPT, 45 CHB and 60 CEU and CEPH populations. We employed a very stringent criterion for the removal of genetic outliers. Two cases with a first principal component of <0.001 and three controls with a fourth principal component of >0.1 were removed from the GWAS analysis (**Supplementary Fig. 1**). The second PCA was used to determine whether population stratification existed between our case and control samples. If strong population stratification existed, the cases and controls would separate into clusters (overlapping or non-overlapping). After removing the population outliers (21 cases and 127 controls), PCA was performed again on the remaining 2,345 samples from the GWAS (746 cases and 1,599 controls). As shown in **Supplementary Figure 1**, there is minimal evidence to suggest that strong population stratification exists in the data. The impact of population stratification was also evaluated by using PCA ($\lambda_{GC} = 1.01$)[9,27].

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 or ~1.01×10^{-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[28]. In general, an I^2 of <30% was considered to signify no heterogeneity, I^2 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[29,30]. All P values from the validation analysis are reported without correction for multiple testing.

Conditional logistic analyses in the GWAS. The three validated SNPs (rs1233710, rs1635 and rs2142731) within 6p21-p22.1 in the current GWAS were in strong LD ($D' > 0.9$ in HapMap CHB and our GWAS samples (**Supplementary Fig. 3**)), which indicated that their associations were probably not independent from each other. Therefore, we selected the top SNP rs1635 (chr. 6: 28.3 Mb, $P_{combined} = 6.91 \times 10^{-12}$) as a marker for association on 6p21-p22.1. Of the top reported SNPs (rs3130375, chr 6: 30.4 Mb, $P_{GWAS} = 3.66 \times 10^{-7}$ from the International Schizophrenia Consortium (ISC); rs3130297, chr. 6: 32.3 Mb, $P_{imputed} = 4.79 \times 10^{-8}$ in the ISC; rs6932590, chr. 6: 27.4 Mb, $P_{combined} = 1.40 \times 10^{-12}$ from the SGENE consortium; rs17619975, chr. 6:15.5 Mb, $P_{GWAS} = 1.49 \times 10^{-5}$ from the Molecular Genetics of Schizophrenia (MGS); and rs13194053, chr. 6: 27.3 Mb, $P_{combined} = 9.5 \times 10^{-9}$ from the ISC and MGS)[6–8], only rs6932590 was analyzed (MAF > 0.01) in the current GWAS, but it did not show any association with schizophrenia ($P_{GWAS} = 0.738$; **Supplementary Table 3**). By searching beyond the top reported SNPs, we found seven previously reported SNPs (including rs6932590 and rs3131296) that were analyzed (MAF > 0.01) in the current GWAS, but only two SNPs rs6913660 and rs6904071 showed moderate associations ($P_{GWAS} = 0.009$) in the current GWAS (**Supplementary Table 3**). Located next to our top SNP is the rs13194053 SNP (chr. 6: 27.3 Mb, $P_{combined} = 9.50 \times 10^{-9}$), which was reported in a meta-analysis performed by the ISC[6]. These two SNPs are not independent ($D' = 1.0$, $r^2 = 1.0$) and showed strong association in individuals of European descent ($P_{combined} = 2.40 \times 10^{-8}$ in the ISC, 1.10×10^{-9} in SGENE and 2.36×10^{-8} in MGS). Therefore, we chose rs6913660 as the marker for schizophrenia association identified in individuals of European descent for our conditional and multi-variable analyses. Then, we performed conditional logistic regression analyses by controlling the allelic dosage for the rs1635 and rs6913660 SNPs as covariates (coded as 0, 1 or 2 alleles), respectively (**Supplementary Table 4**).

25. Chen, J. *et al.* Genetic structure of the Han Chinese population revealed by genome-wide SNP variation. *Am. J. Hum. Genet.* **85**, 775–785 (2009).
26. Xu, S. *et al.* Genomic dissection of population substructure of Han Chinese and its implication in association studies. *Am. J. Hum. Genet.* **85**, 762–774 (2009).
27. Zhang, F.R. *et al.* Genomewide association study of leprosy. *N. Engl. J. Med.* **361**, 2609–2618 (2009).
28. Higgins, J.P. & Thompson, S.G. Quantifying heterogeneity in a meta-analysis. *Stat. Med.* **21**, 1539–1558 (2002).
29. Higgins, J.P. *et al.* Measuring inconsistency in meta-analyses. *Br. Med. J.* **327**, 557–560 (2003).
30. DerSimonian, R. & Laird, N. Meta-analysis in clinical trials. *Control. Clin. Trials* **7**, 177–188 (1986).