Common variants on 8p12 and 1q24.2 confer risk of schizophrenia

Yongyong Shi[1–3,32], Zhiqiang Li[1,2,32], Qi Xu[4,32], Ti Wang[1], Tao Li[1], Jiawei Shen[1], Fengyu Zhang[5], Jianhua Chen[6], Guoquan Zhou[3], Weidong Ji[3], Baojie Li[1], Yifeng Xu[6], Dengtang Liu[6], Peng Wang[7], Ping Yang[7], Benxiu Liu[8], Wensheng Sun[8], Chunling Wan[1], Shengying Qin[1,2], Guang He[1,2], Stacy Steinberg[9], Sven Cichon[10–12], Thomas Werge[13], Engilbert Sigurdsson[14], Sarah Tosato[15], Aarno Palotie[16–20], Markus M Nöthen[10,11,21], Marcella Rietschel[22,23], Roel A Ophoff[24–26], David A Collier[27], Dan Rujescu[28], David St Clair[29], Hreinn Stefansson[9], Kari Stefansson[9], Jue Ji[1], Qingzhong Wang[1], Wenjin Li[1], Linqing Zheng[1], Hairong Zhang[1], Guoyin Feng[5] & Lin He[1,2,30,31]

Schizophrenia is a severe mental disorder affecting ~1% of the world population, with heritability of up to 80%. To identify new common genetic risk factors, we performed a genome-wide association study (GWAS) in the Han Chinese population. The discovery sample set consisted of 3,750 individuals with schizophrenia and 6,468 healthy controls (1,578 cases and 1,592 controls from northern Han Chinese, 1,238 cases and 2,856 controls from central Han Chinese, and 934 cases and 2,020 controls from the southern Han Chinese). We further analyzed the strongest association signals in an additional independent cohort of 4,383 cases and 4,539 controls from the Han Chinese population. Meta-analysis identified common SNPs that associated with schizophrenia with genome-wide significance on 8p12 (rs16887244, $P = 1.27 \times 10^{-10}$) and 1q24.2 (rs10489202, $P = 9.50 \times 10^{-9}$). Our findings provide new insights into the pathogenesis of schizophrenia.

Schizophrenia is a severe neuropsychiatric disorder characterized by psychotic behavior (delusion and hallucinations), disorganization, dysfunction in normal affective responses and altered cognitive functioning[1]. Recent genome-wide studies of populations of European ancestry have revealed both common, low-risk SNPs and rare, high-penetrance copy-number variants predisposing individuals to schizophrenia[2–8]. However, much of the heritability of schizophrenia remains unaccounted for, and the pathways or biological mechanisms that underlie susceptibility are still largely unknown.

To search for common variants associated with schizophrenia, we performed a GWAS of 3,750 individuals with schizophrenia and 6,468 healthy controls genotyped using Affymetrix SNP 6.0 GeneChips in Han Chinese populations from three geographic locations (northern, central and southern China). In total, 546,561 SNPs were used for statistical analysis after quality control filtering in the initial studies. Logistic regression was used to test the additive effects of minor allele dosage for each SNP. Potential population stratification was adjusted for using eigenvectors from principal-component analysis (PCA) (see Online Methods).

We carried out a genome-wide combined analysis in three independent cohorts in the discovery stage and then further analyzed the most significantly associated SNPs in additional samples. Initially, we performed a GWAS that included (i) 1,578 subjects with schizophrenia

Table 1 GWAS, replication study and meta-analysis results for selected SNPs

| | | | | | BIOX GWAS | | | | | | | | BIOX replication (4,383 cases, 4,539 controls) | | | BIOX GWAS and replication |
| | | | | | Northern Han (1,578 cases, 1,592 controls) | | Central Han (1,238 cases, 2,856 controls) | | Southern Han (934 cases, 2,020 controls) | | Initial meta | | | | | |
Chr.	SNP	Location (bp)	Allele[b]	Frequency[a]	OR (95% CI)	P^c	OR (95% CI)	P^c	OR (95% CI)	P^c	$P_{GWAS-meta}$	Allele	OR (95% CI)	P_{REP}	$P_{BIOX-meta}$
1	rs10489202	166169703	A	0.141	1.273 (1.084–1.494)	0.003	1.309 (1.122–1.528)	6.19×10^{-4}	1.180 (0.953–1.461)	0.130	2.65×10^{-6}	A	1.189 (1.094–1.293)	4.76×10^{-5}	9.50×10^{-9}
1	rs1060041	166240600	T	0.139	1.272 (1.083–1.493)	0.003	1.324 (1.133–1.547)	4.06×10^{-4}	1.136 (0.916–1.410)	0.245	4.11×10^{-6}	T	1.113 (1.025–1.211)	0.012	5.31×10^{-7}
1	rs11586522	166357655	T	0.160	1.277 (1.099–1.483)	0.001	1.267 (1.101–1.459)	9.77×10^{-4}	1.181 (0.956–1.46)	0.123	1.53×10^{-6}	T	1.040 (0.957–1.129)	0.362	1.17×10^{-4}
8	rs16887244	38150502	G	0.317	0.852 (0.747–0.972)	0.017	0.809 (0.718–0.912)	5.16×10^{-4}	0.876 (0.745–1.029)	0.107	9.27×10^{-6}	G	0.833 (0.781–0.888)	2.23×10^{-8}	1.27×10^{-10}
8	rs1488935	38252950	T	0.316	0.842 (0.737–0.962)	0.011	0.795 (0.703–0.898)	2.30×10^{-4}	0.871 (0.742–1.023)	0.093	2.81×10^{-6}	T	0.871 (0.816–0.929)	3.42×10^{-5}	5.06×10^{-9}

aFrequency is the average minor allelic frequency of controls in the BIOX data. bMinor allele. cThese P values were adjusted by PCA methods in the GWAS of each geographical group.

and 1,592 healthy controls from the northern Han Chinese population (Beijing and Shandong provinces), (ii) 1,238 subjects with schizophrenia and 2,856 healthy controls from the central Han Chinese population (Shanghai and Anhui provinces), and (iii) 934 subjects with schizophrenia and 2,020 healthy controls from the southern Han Chinese population (Guangdong and Guangxi provinces). Sample descriptions and characteristics can be found in Online Methods and **Supplementary Table 1**. We carried out a meta-analysis using the PCA-adjusted association results of the three cohorts (see Online Methods and **Supplementary Fig. 1**). For all reported SNPs, the observed P values were identical when the fixed- and random-effects models were used for meta-analysis, and no heterogeneity was found among the three cohorts (**Supplementary Table 2**). Although no SNP reached genome-wide significance ($P < 5.0 \times 10^{-8}$), four SNPs on 8p12 (rs1488935, $P_{GWAS-meta} = 2.81 \times 10^{-6}$) and 1q24.2 (rs10489202, $P_{GWAS-meta} = 2.65 \times 10^{-6}$; rs1060041, $P_{GWAS-meta} = 4.11 \times 10^{-6}$; rs11586522, $P_{GWAS-meta} = 1.53 \times 10^{-6}$) showed strong association ($P_{GWAS-meta} < 5.0 \times 10^{-6}$) (**Table 1**, **Supplementary Fig. 2** and **Supplementary Table 2**). We examined the top SNPs in these regions (three SNPs on 1q24.2 and two on 8p12) for association in an independent cohort of 4,383 subjects with schizophrenia and 4,539 healthy controls, which, together with the discovery sample set, we defined as the BIOX sample. We successfully replicated the association findings for 8p12 and 1q24.2, and the associations in the BIOX sample reached genome-wide significance (8p12: rs16887244, $P_{BIOX-meta} = 1.27 \times 10^{-10}$, G allele odds ratio (OR) = 0.84, 95% confidence interval (CI) = 0.79–0.88; rs1488935, $P_{BIOX-meta} = 5.06 \times 10^{-9}$, T allele OR = 0.85, 95% CI = 0.81–0.90; 1q24.2: rs10489202, $P_{BIOX-meta} = 9.50 \times 10^{-9}$, A allele OR = 1.23, 95% CI = 1.15–1.32) (**Table 1**). To search for additional associations of common variants, we also carried out imputation analysis in the selected regions (**Fig. 1**).

We further validated our top association signals by obtaining results from the SGENE-plus project, a large-scale genome-wide association study of schizophrenia that aims to identify genetic variants associated with the disease and to study their impact on phenotype and their interaction with environmental factors contributing to pathogenesis. This GWAS included 3,830 subjects with schizophrenia and 14,724 controls of European ancestry. Both rs16887244 ($P_{SGENE-plus} = 0.026$, G allele OR = 0.92, 95% CI = 0.85–0.99) and rs1488935 ($P_{SGENE-plus} = 0.027$, T allele OR = 0.92, 95% CI = 0.85–0.99) showed nominal association with schizophrenia in the population of European ancestry, and the direction of the effect was consistent with our findings (**Supplementary Tables 2** and **3**). The effect size for rs10489202 (A allele OR = 1.01, 95% CI = 0.94–1.09) was small in the SGENE-plus data (**Supplementary Table 3**). A statistical power comparison between the two data sets was also carried out (**Supplementary Table 4**).

On 8p12, both the rs16887244 and rs1488935 SNPs reached genome-wide significance in the combined analysis (**Fig. 1a** and **Table 1**). Controlling for rs16887244, conditional logistic regression analysis revealed that there were no additional association signals (**Supplementary Table 5**). The rs16887244 SNP is located in intron 1 of *LSM1* (MIM: 607281), and rs1488935 is located in intron 23 of *WHSC1L1* (MIM: 607083). We used data from two published expression quantitative trait loci (eQTL) data sets (derived from lymphoblastoid cell lines[9,10] and the brain[11]) to determine whether rs16887244 is associated with expression of the nearby genes. In the expression data from the lymphoblastoid cell lines, rs16887244 is nominally associated with the expression of several genes (*ASH2L*, *LSM1*, *BAG4*, *DDHD2* and *PPAPDC1B*; $P < 0.05$) (**Supplementary Table 6**), and in the expression data from the brain, it is nominally associated with the expression of the *ASH2L*, *DDHD2*, *PPAPDC1B* and *LETM2* genes ($P < 0.05$) (**Supplementary Table 7**).
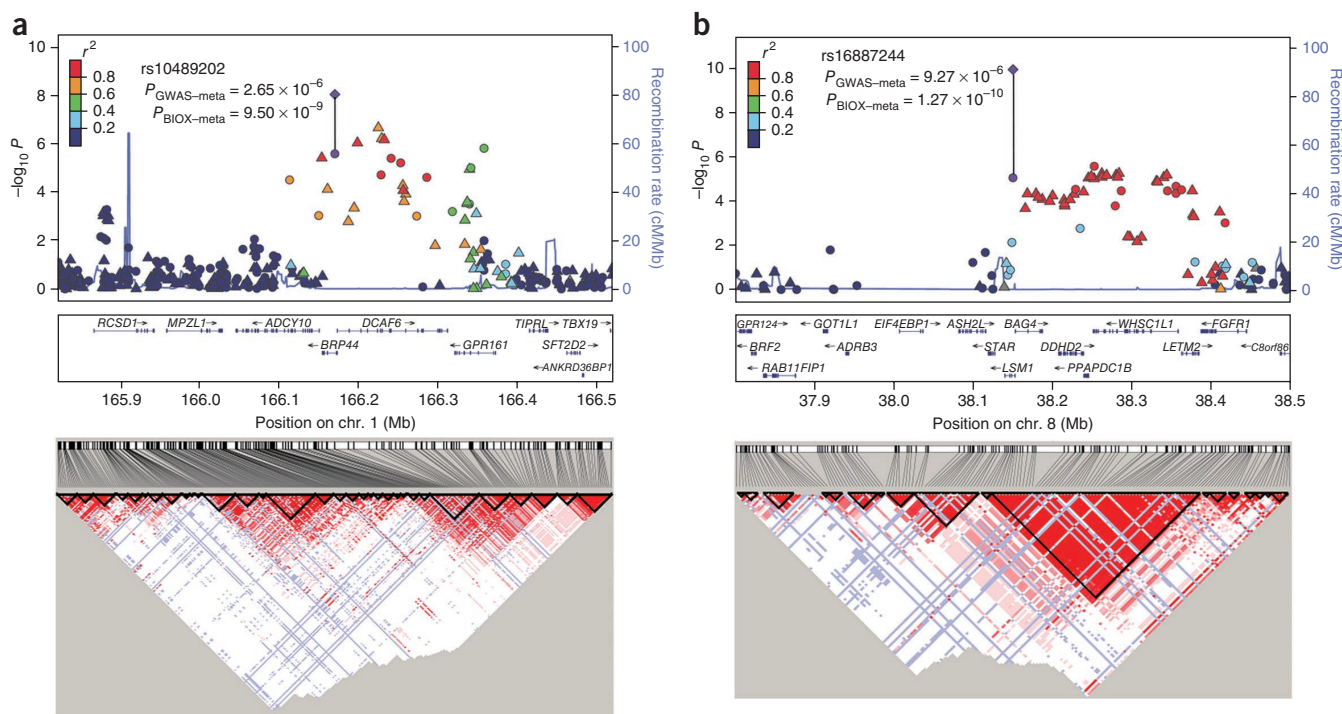
Figure 1 Regional plots of the two loci associated with schizophrenia. (**a,b**) Results are shown for the 8p12 (**a**) and 1q24.2 (**b**) regions. Top, $-\log_{10}$ P values are shown for SNPs for the region 350 kb on either side of the marker SNPs. $P_{\text{GWAS–meta}}$ is for results obtained from the combination of the three initial data sets and is shown for genotyped (circle) and imputed (triangle) SNPs. $P_{\text{BIOX–meta}}$ is for results obtained from the combination of the initial and replication study data in the BIOX data set (diamond). The marker SNP is shown in purple, and the r^2 values of the other SNPs are indicated by color. The genes within the relevant regions are annotated and shown as arrows. Bottom, the LD block structure of the related surrounding region is shown according to CHB and Japanese in Tokyo (JPT) HapMap data.

Notably, rs1488935 is located ~135 kb upstream of *FGFR1* (MIM: 136350), encoding the fibroblast growth factor receptor 1 protein. Hippocampal *FGFR1* mRNA expression has been reported to be upregulated in schizophrenia and in individuals with major depression[12], and transgenic mice expressing a dominant-negative mutant (FGFR1(TK−)) from the catecholaminergic, neuron–specific tyrosine hydroxylase (*Th*) gene promoter have been found to show a schizophrenia-like syndrome[13]. Several lines of evidence support a role for fibroblast growth factors (FGFs) in schizophrenia, including functional plausibility, positional and functional genetic studies, knockout mouse models, examination of the effects of FGF in animals and humans, and the study of the association between FGFs and environmental risk factors for schizophrenia[14]. Moreover, it has been reported that another FGF receptor gene, *FGFR2,* is associated with schizophrenia ($P = 0.0009$)[15].

On 1q24.2, rs10489202 is located in intron 1 of *BRP44*, encoding brain protein 44 (**Fig. 1b**), rs1060041 is a coding-synonymous SNP in *DCAF6* (MIM: 610494) and rs11586522 is located in the intron 2 region of *GPR161* (MIM: 612250). Controlling for the most significant signal from rs10489202, logistic regression analysis indicated that there was no additional signal in the genomic region surrounding rs10489202 (**Supplementary Table 5**). In expression data from the lymphoblastoid cell lines, rs10489202 is nominally associated with the expression of several genes (*MPZL1, DCAF6* and *TIPRL*; $P < 0.05$) (**Supplementary Table 6**); however, in expression data from the brain, no association was observed (**Supplementary Table 7**). Of note, rs10489202 is ~140 kb downstream of *MPZL1* (MIM: 604376), which encodes the myelin protein zero–like 1 or protein zero–related. A study found that *MPZL1* (PZR) was significantly upregulated in

individuals with schizophrenia relative to healthy controls (1.29 fold, $P = 0.0263$)[16]. Common SNPs in *MPZL1* were recently reported to be associated with schizophrenia in a case-parent triad study of 523 subjects in the Han Chinese population ($P = 0.0017$)[17].

Population stratification is a major concern in GWAS that use the case-control design. It has been demonstrated using a group of GWAS data sets that there is a "north-south" population structure in China[18]. PCA analysis of all initial study samples also revealed such a structure (**Supplementary Figs. 3** and **4**). We therefore stratified our samples into northern, central and southern groups according to their geographic region and carried out GWAS analysis on the individual groups. Combined association results were then obtained using meta-analysis.

The major histocompatibility (MHC) region on 6p21.3–22.1 has previously been associated with schizophrenia in GWAS of individuals of European ancestry[4–6], and these findings have been replicated in a large sample of the Han Chinese population[19]. In our initial study, no common SNP in this region met the selection criteria ($P_{\text{GWASs–meta}} < 5.0 \times 10^{-6}$). However, 149 SNPs out of 1,786 (8.34%, $P_{\text{GWAS–meta}} < 0.05$) showed nominal association with schizophrenia in the MHC region (**Supplementary Table 8**), and the most significant signal was observed at rs2394514 ($P_{\text{GWASs–meta}} = 1.16 \times 10^{-5}$), which is located at 6p22.1. We found no significant association for rs13194053 (refs. 4,5) and rs6932590 (ref. 6) on 6p22.1 and rs3131296 (ref. 6) on 6p21.32, which were previously reported to have associations with schizophrenia of genome-wide significance ($P < 5.0 \times 10^{-8}$) in populations of European ancestry (derived from the National Human Genome Research Institute GWAS catalog[20]; a full listing of SNPs is provided in **Supplementary Table 9**). We compared the allele frequencies of these

SNPs in samples from the Utah residents of Northern and Western European descent (CEU) and Chinese Han in Beijing (CHB) HapMap groups (**Supplementary Table 10**). The allele frequencies for most of the SNPs were noticeably different between these European- and Chinese-descended populations. In addition, we measured linkage disequilibrium (LD; r^2) values for these SNPs with the 149 SNPs (in the MHC region, with $P_{GWASs-meta} < 0.05$ in our study) in the HapMap CEU and HCB samples (**Supplementary Tables 11 and 12**). The rs7749823 SNP, which is in a LD block with rs13194053 ($r^2 = 0.644$) and rs6932590 ($r^2 = 0.623$) in populations of European ancestry, showed association with schizophrenia in our study ($P_{GWASs-meta} = 4.18 \times 10^{-5}$, A allele OR = 1.64, risk allele frequency (RAF) = 0.97). Haplotype analysis of the HapMap CEU sample showed the frequency of the TTA haplotype at rs13194053, rs6932590 and rs7749823 to be 0.808. As both the T alleles of rs13194053 and rs6932590 are risk alleles for schizophrenia in the European population, our finding of association for the A allele of rs7749823 in the three-marker haplotype in the Han Chinese population is consistent with the findings in individuals of European descent.

We also examined other known schizophrenia susceptibility loci (*NRGN* on 11q24.2 and *TCF4* on 18q21.2) that were reported to have genome-wide significance in published studies[4–6] (**Supplementary Table 9**). None of the SNPs within these loci show association with schizophrenia in our data set. However, genetic heterogeneity obviously exists for schizophrenia risk variants across different ancestries (**Supplementary Table 10**). The presence of this heterogeneity reinforces the need for GWAS of schizophrenia to be carried out in diverse populations.

In summary, we have identified common genetic variants on 8p12 and 1q24.2 that are associated with schizophrenia in the Han Chinese population. Several promising candidate genes are implicated in these two regions, making it difficult to determine precisely which genes contain the causative variants. Nevertheless, the identification of these new common genetic risk variants that predispose individuals to schizophrenia is an encouraging first step in a process that has the potential to translate into improved methods for the prediction and treatment of this disease.

URLs. R, http://www.r-project.org/; LocusZoom, http://csg.sph.umich.edu/locuszoom/; EIGENSTRAT, http://genepath.med.harvard.edu/~reich/Software.htm; PLINK, http://pngu.mgh.harvard.edu/~purcell/plink/; Haploview program, http://www.broad.mit.edu/mpg/haploview; The International HapMap Project, http://hapmap.ncbi.nlm.nih.gov/; SHEsis, http://analysis.bio-x.cn/myAnalysis.php; MACH, http://www.sph.umich.edu/csg/abecasis/MACH/index.html; NCBI GEO, http://www.ncbi.nlm.nih.gov/geo/.

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.

AUTHOR CONTRIBUTIONS

Y.S. and L.H. conceived of and designed the study. Y.S. supervised all the experiments and data analysis. Y.S. and Z.L. conducted data analyses and drafted the manuscript. Y.S., Z.L., F.Z., D.S.C., S.S., D.R. and L.H. revised the manuscript. Y.S., G.F., Q.X., J.C., Y.X., D.L., P.W., P.Y., B. Liu, W.S., G.Z. and W.J. recruited samples. T. Wang, J.J., T.L., J.S., J.C., Q.W., W.L., L.Z., H.Z., B. Li, C.W., S.Q. and G.H. performed or contributed to the experiments. S.S., S.C., T.W., E.S., S.T., A.P., M.M.N., M.R., R.A.O., D.A.C., D.R., D.S.C., H.S. and K.S. provided the SGENE-plus data. All authors critically reviewed and approved the manuscript.

1. Craddock, N., O'donovan, M.C. & Owen, M.J. The genetics of schizophrenia and bipolar disorder: dissecting psychosis. *J. Med. Genet.* **42**, 193–204 (2005).
2. McCarthy, S.E. *et al.* Microduplications of 16p11. 2 are associated with schizophrenia. *Nat. Genet.* **41**, 1223–1227 (2009).
3. 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).
4. Purcell, S.M. *et al.* Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748–752 (2009).
5. Shi, J. *et al.* Common variants on chromosome 6p22. 1 are associated with schizophrenia. *Nature* **460**, 753–757 (2009).
6. Stefansson, H. *et al.* Common variants conferring risk of schizophrenia. *Nature* **460**, 744–747 (2009).
7. Vacic, V. *et al.* Duplications of the neuropeptide receptor gene *VIPR2* confer significant risk for schizophrenia. *Nature* **471**, 499–503 (2011).
8. Xu, B. *et al.* Strong association of *de novo* copy number mutations with sporadic schizophrenia. *Nat. Genet.* **40**, 880–885 (2008).
9. Stranger, B.E. *et al.* Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* **315**, 848–853 (2007).
10. Stranger, B.E. *et al.* Population genomics of human gene expression. *Nat. Genet.* **39**, 1217–1224 (2007).
11. Gibbs, J.R. *et al.* Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. *PLoS Genet.* **6**, e1000952 (2010).
12. Gaughran, F., Payne, J., Sedgwick, P.M., Cotter, D. & Berry, M. Hippocampal FGF-2 and FGFR1 mRNA expression in major depression, schizophrenia and bipolar disorder. *Brain Res. Bull.* **70**, 221–227 (2006).
13. Klejbor, I. *et al.* Fibroblast growth factor receptor signaling affects development and function of dopamine neurons–inhibition results in a schizophrenia-like syndrome in transgenic mice. *J. Neurochem.* **97**, 1243–1258 (2006).
14. Terwisscha van Scheltinga, A.F., Bakker, S.C. & Kahn, R.S. Fibroblast growth factors in schizophrenia. *Schizophr. Bull.* **36**, 1157–1166 (2010).
15. O'Donovan, M.C. *et al.* Analysis of 10 independent samples provides evidence for association between schizophrenia and a SNP flanking fibroblast growth factor receptor 2. *Mol. Psychiatry* **14**, 30–36 (2009).
16. Tkachev, D. *et al.* Oligodendrocyte dysfunction in schizophrenia and bipolar disorder. *Lancet* **362**, 798–805 (2003).
17. He, G. *et al.* *MPZL1/PZR*, a novel candidate predisposing schizophrenia in Han Chinese. *Mol. Psychiatry* **11**, 748–751 (2006).
18. 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).
19. Li, T. *et al.* Common variants in major histocompatibility complex region and *TCF4* gene are significantly associated with schizophrenia in Han Chinese. *Biol. Psychiatry* **68**, 671–673 (2010).
20. Hindorff, L.A. *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. USA* **106**, 9362–9367 (2009).

ONLINE METHODS

Subjects. All individuals with schizophrenia analyzed in the BIOX study were interviewed by two independent psychiatrists, were diagnosed according to Diagnostic and Statistical Manual of Mental Disorders (DSM)-IV criteria, and had a 2-year history of the disease. All cases met two criteria—preoccupation with one or more delusions and frequent auditory hallucinations—but were not strongly affected with any of the following symptoms: disorganized speech, disorganized or catatonic behavior, or flat or inappropriate actions. The schizophrenia cases in the initial and replication stages of our study were recruited using the same diagnostic criteria.

Healthy controls were randomly selected from Han Chinese volunteers who were requested to reply to a written invitation to evaluate their medical history. Lists of potential control subjects were screened for suitable volunteers by excluding subjects with major mental illness.

The discovery phase (BIOX GWAS) included three data sets: the northern Han Chinese set of 1,578 cases and 1,592 controls recruited from Beijing and Shandong provinces, the central Han Chinese set of 1,238 cases and 2,856 controls recruited from Shanghai and Anhui provinces, and the southern Han Chinese set of 934 cases and 2,020 controls recruited from Guangdong and Guangxi provinces. Sample descriptions can be found in **Supplementary Table 1**.

The replication stage (BIOX replication) included 4,383 schizophrenia cases and 4,539 healthy controls recruited from Shanghai. All subjects in the replication stage were unrelated, were born in Shanghai and had parents who were also residents of Shanghai. This replication sample was expected to have minimal population stratification, given that Shanghai used to have the strictest resident registration system in China before the 1980s. Sample descriptions and characteristics can be found in **Supplementary Table 1**.

The SGENE-plus samples included 513 schizophrenia cases and 471 controls from Denmark, 93 cases and 88 unrelated controls from England, 182 cases and 197 controls from Finland, 1,048 cases and 971 controls from Germany, 531 cases and 11,615 controls from Iceland, 84 cases and 89 controls from Italy, 693 cases and 629 controls from The Netherlands, and 658 cases and 661 controls from Scotland. All cases were diagnosed according to International Statistical Classification of Diseases (ICD-10) or DSM-IV criteria. These samples were genotyped and used in a previously published GWAS[6] with two exceptions: (i) slight changes in the Icelandic schizophrenia cases used (~10% difference) and (ii) two newly genotyped genome-wide sample sets from The Netherlands and Denmark.

All participants provided written informed consent. Approval was received for our study from the local Ethics Committee of Human Genetic Resources. Other details of sample description are provided in the **Supplementary Note**.

DNA extraction. Venous blood samples anti-coagulated with EDTA were collected from all participants. Genomic DNA was extracted from peripheral blood lymphocytes by standard procedures using FlexiGene DNA kits (Fuji) and diluted to working concentrations of 50 ng/ml for SNP chip genotyping and 10–20 ng/ml for replication genotyping.

GWAS genotyping and quality control. The genome-wide association analysis was performed using the Affymetrix Genome-Wide Human SNP Array 6.0. Quality control (QC) filtering of the GWAS data was performed by excluding arrays with Contrast QC <0.4 from further data analysis. Genotype data were generated using the birdseed algorithm[21]. For sample filtering, arrays that generated genotypes at <95% of loci were excluded. For SNP filtering (after sample filtering), SNPs with call rates <95% in either cases or controls were removed in each geographic group. SNPs with minor allele frequency (MAF) <1% or significant deviation from Hardy-Weinberg equilibrium (HWE; $P \leq 1 \times 10^{-3}$) in controls were also excluded. SNPs passing QC standards in the northern, central and southern Han Chinese groups were used for further analysis. After QC filtering, there were 546,561 SNPs remaining in the combined initial study.

SNP selection criteria and replication genotyping. We selected all SNPs with $P_{GWAS-meta} < 5.0 \times 10^{-6}$ for the replication study, according to a previous GWAS[22]. To ensure that there were at least two SNPs included for each region, we added the second most significant SNP on 8p12 for replication. Genotyping for

the replication study was performed using the ligation detection reaction (LDR) method[23,24], with technical support from the Shanghai Biowing Applied Biotechnology company.

Analysis of population substructure. Population substructure was evaluated using principal-components analysis (PCA) using EIGENSTRAT software[25,26]. Twenty components, some of which were predicted to reflect ancestry differences among subjects, were generated for each sample. Logistic regression was used to determine whether there was a significant difference in component scores between cases and controls; significant components were used as covariates in the association analysis to correct for population stratification. We identified six significant components for the northern Han Chinese data set, eight for the central Han Chinese data set and four for the southern Han Chinese data set.

Imputation analysis of ungenotyped SNPs. We selected the two newly identified potential susceptibility loci on 1q24.2 (rs10489202 and 350 kb up- and downstream) and on 8p12 (rs16887244 and 350 kb up- and downstream) for imputation. HapMap SNPs in the two regions were imputed using MACH 1.0. Phased haplotypes for 90 CHB and JPT subjects (180 haplotypes) were used as the reference for imputing genotypes. Any SNP imputed with information content of $r^2 < 0.3$ was excluded from association analysis because of lack of power. The criteria for imputed SNP quality control filtering were the same as for the genotyped SNPs.

Statistical methods. The association of single SNPs with schizophrenia was tested by logistic regression using PLINK[27] that was performed separately for the northern, central and southern Han Chinese data sets, correcting for principal-component scores that had statistically significant differences between cases and controls. HWE analysis was performed using PLINK, and Haploview[28] was used to generate genome-wide P plots. Quantile-quantile plots were created using the R package, and regional plots were generated using LocusZoom (see URLs). In the replication study, allelic association analysis was conducted using SHEsis[29]. A meta-analysis using the random-effect model was carried out on the basis of the PCA-adjusted association results of the three cohorts in the initial GWAS using PLINK, and the inflation value was 0.97. Heterogeneity across the data sets was evaluated using the Cochran's Q test. The meta-analysis was carried out using the Mantel-Haenszel method with a random-effects model[30]. Conditional logistic regression was used to test for the independent effects of an individual SNP[22,31]. We compared the statistical power between BIOX and SGENE-plus on the basis of the sample size of the BIOX and SGENE-plus data sets ($\alpha = 5 \times 10^{-8}$; **Supplementary Table 4**)[5,32].

Expression quantitative trait loci (eQTL) analysis. Expression profiles were analyzed within two eQTL data sets (lymphoblastoid cell lines and the brain). We downloaded the expression data sets from the NCBI GEO database. The first data set[9,10] consists of gene expression profiles generated using RNA extracted from lymphoblastoid cell lines derived from the 210 unrelated HapMap individuals from four sample groups (60 CEU, 45 CHB, 45 JPT and 60 Yoruba in Ibadan (YRI)). Expression analysis was performed using Sentrix Human-6 Expression BeadChips (Illumina). The SNP genotypes from HapMap 2 were used in the analysis. The second eQTL data set[11] consists of gene-expression profiles for frozen tissue samples obtained from four brain regions (the cerebellum, pons and frontal and temporal cortices) of 143 neurologically normal subjects of European ancestry. Expression analysis was performed using Illumina HumanRef-8 Expression BeadChips, and SNP genotyping was carried out using Infinium HumanHap550 BeadChips. The expression data were normalized and log transformed, as described in the original reports[9–11]. The eQTLs were tested by linear regression of normalized expression levels on SNP genotypes (coded as the number of minor alleles at each SNP: 0, 1 or 2). For the lymphoblastoid cell line data set, analyses were conducted for each population and the combined data set. For the brain data set, analysis of each tissue region was performed separately.

21. Korn, J.M. *et al.* Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. *Nat. Genet.* **40**, 1253–1260 (2008).

22. Chen, Z.J. *et al.* Genome-wide association study identifies susceptibility loci for polycystic ovary syndrome on chromosome 2p16.3, 2p21 and 9q33.3. *Nat. Genet.* **43**, 55–59 (2011).
23. Thomas, G. *et al.* Capillary and microelectrophoretic separations of ligase detection reaction products produced from low-abundant point mutations in genomic DNA. *Electrophoresis* **25**, 1668–1677 (2004).
24. Yi, P. *et al.* PCR/LDR/capillary electrophoresis for detection of single-nucleotide differences between fetal and maternal DNA in maternal plasma. *Prenat. Diagn.* **29**, 217–222 (2009).
25. Patterson, N., Price, A.L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2**, e190 (2006).
26. Price, A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
27. 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).
28. Barrett, J.C., Fry, B., Maller, J. & Daly, M.J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263–265 (2005).
29. Shi, Y.Y. & He, L. SHEsis, a powerful software platform for analyses of linkage disequilibrium, haplotype construction, and genetic association at polymorphism loci. *Cell Res.* **15**, 97–98 (2005).
30. Mantel, N. & Haenszel, W. Statistical aspects of the analysis of data from retrospective studies. *J. Natl. Cancer Inst.* **22**, 719–748 (1959).
31. Petukhova, L. *et al.* Genome-wide association study in alopecia areata implicates both innate and adaptive immunity. *Nature* **466**, 113–117 (2010).
32. Menashe, I., Rosenberg, P.S. & Chen, B.E. PGA: power calculator for case-control genetic association analyses. *BMC Genet.* **9**, 36 (2008).