

Published in final edited form as:

*Nat Genet.* ; 43(12): 1224–1227. doi:10.1038/ng.980.

## Common Variants on 8p12 and 1q24.2 Confer Risk of Schizophrenia

Yongyong Shi<sup>1,2,#,\*</sup>, Zhiqiang Li<sup>1,#</sup>, Qi Xu<sup>3,#</sup>, Ti Wang<sup>1</sup>, Tao Li<sup>1</sup>, Jiawei Shen<sup>1</sup>, Fengyu Zhang<sup>4</sup>, Jianhua Chen<sup>5</sup>, Guoquan Zhou<sup>2</sup>, Weidong Ji<sup>2</sup>, Baojie Li<sup>1</sup>, Yifeng Xu<sup>5</sup>, Dengtang Liu<sup>5</sup>, Peng Wang<sup>6</sup>, Ping Yang<sup>6</sup>, Benxiu Liu<sup>7</sup>, Wensheng Sun<sup>7</sup>, Chunling Wan<sup>1</sup>, Shengying Qin<sup>1</sup>, Guang He<sup>1</sup>, Stacy Steinberg<sup>8</sup>, Sven Cichon<sup>9,10</sup>, Thomas Werge<sup>11</sup>, Engilbert Sigurdsson<sup>12</sup>, Sarah Tosato<sup>13</sup>, Aarno Palotie<sup>14,15,16,17</sup>, Markus M. Nöthen<sup>9</sup>, Marcella Rietschel<sup>18,19</sup>, Roel A. Ophoff<sup>20,21</sup>, David A. Collier<sup>22</sup>, Dan Rujescu<sup>23</sup>, David St Clair<sup>24</sup>, Hreinn Stefansson<sup>8</sup>, Kari Stefansson<sup>8</sup>, Jue Ji<sup>1</sup>, Qingzhong Wang<sup>1</sup>, Wenjin Li<sup>1</sup>, Linqing Zheng<sup>1</sup>, Hairong Zhang<sup>1</sup>, Guoyin Feng<sup>5</sup>, and Lin He<sup>1,25,26,\*</sup>

<sup>1</sup>Bio-X Institutes, Key Laboratory for the Genetics of Developmental and Neuropsychiatric Disorders (Ministry of Education), Shanghai Jiao Tong University, Shanghai 200030, P.R. China <sup>2</sup>Changning Mental Health Center, Shanghai 200335, P.R. China <sup>3</sup>National Laboratory of Medical Molecular Biology, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100005, P.R. China <sup>4</sup>Clinical Brain Disorders Branch, GCAP, IRP, NIMH, Bethesda, MD, 20892 <sup>5</sup>Shanghai Institute of Mental Health, Shanghai 200030, P.R. China <sup>6</sup>Fourth People's Hospital, Wuhu 241003, Anhui, P.R. China <sup>7</sup>Longquan Mountain Hospital of Guangxi Province, 1 Jila Road, Liuzhou 545005, P.R. China <sup>8</sup>deCODE genetics, Sturlugata 8, IS-101, Reykjavik, Iceland <sup>9</sup>Institute of Human Genetics Department of Genomics Life & Brain Center University of Bonn Sigmund-Freud-Strasse 25 D-53105 Bonn, Germany <sup>10</sup>Institute of Neuroscience and Medicine (INM-1) Structural and Functional Organization of the Brain Genomic Imaging Research Center Juelich Leo-Brandt-Strasse 5 D-52425 Juelich <sup>11</sup>Institute of Biological Psychiatry, Mental Health Centre Sct. Hans Copenhagen University Hospital, DK-4000 Roskilde, Denmark <sup>12</sup>Department of Psychiatry, National University

\*Correspondence should be addressed to: Yongyong Shi or Lin He, Bio-X Institutes, Shanghai Jiao Tong University, the Little White House, 1954 Huashan Road, Shanghai 200030, P.R.China or Institute of Neuropsychiatric Science and Systems Biological Medicine, Changning Mental Health Center, the Bio-X Center Hospital, 299 XieHe Road, Shanghai 200042, P.R.China or Institutes of Biomedical Sciences, Fudan University, 130 Dongan Road, Shanghai 200032, China. Tel & Fax: +86 21 62933338. shiyongyong@gmail.com or helinhe@gmail.com.

#These authors are co-first authors of this study.

### Author Contributions

Y.S. and L.H. conceived and designed the study. Y.S. supervised all the experiments and data analysis. Y.S., 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.X.L., W.S., G.Z., W.J. recruited samples. Ti.W., J.J., T.L., J.S., J.C., Q.W., W.L., L.Z., H.Z., B.J.L., C.W., S.Q., 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., K.S. provided the SGENE-plus data. All authors critically reviewed and approved the manuscript.

### Competing Financial Interests

The authors declare no competing financial interests.

### 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://www.hapmap.org>

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/>

Hospital, Hringbraut, IS-101 Reykjavik, Iceland <sup>13</sup>Section of Psychiatry and Clinical Psychology, University of Verona, Verona, 37134 Verona, Italy <sup>14</sup>Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 <sup>15</sup>Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland <sup>16</sup>Program in Medical and Population Genetics and Genetic Analysis Platform, The Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA <sup>17</sup>Department of Medical Genetics, University of Helsinki and University Central Hospital, Helsinki, Finland <sup>18</sup>Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, University of Heidelberg, J5, D-68159 Mannheim, Germany <sup>19</sup>Department of Psychiatry, University of Bonn, Sigmund-Freud-Strasse 25 D-53105 Bonn, Germany <sup>20</sup>Rudolf Magnus Institute of Neuroscience and Department of Medical Genetics, University Medical Center, 3584 CG Utrecht, The Netherlands <sup>21</sup>Center for Neurobehavioral Genetics, UCLA, Los Angeles, CA 90095, USA <sup>22</sup>Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, King's College, London SE5 8AF, UK <sup>23</sup>Division of Molecular and Clinical Neurobiology, Department of Psychiatry, Ludwig-Maximilians-University, Nußbaumstrasse 7, 80336 Munich, Germany <sup>24</sup>Department of Mental Health, University of Aberdeen, Royal Cornhill Hospital, Aberdeen AB25 2ZD, UK <sup>25</sup>Institutes of Biomedical Sciences, Fudan University, Shanghai 200032, PR China <sup>26</sup>Institute for Nutritional Sciences, Shanghai Institute of Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, P.R. China

## Abstract

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 patients and 6,468 healthy controls (1,578 cases and 1,592 controls from the Northern Han; 1,238 cases and 2,856 controls from the Central Han; 934 cases and 2,020 controls from the Southern Han); and we followed up the top association signals in an additional independent cohort of 4,383 cases and 4,539 controls from the Han Chinese. Meta-analysis identified genome-wide significant association of common SNPs with schizophrenia on chromosome 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 features (delusions and hallucinations), disorganization, dysfunction in normal affective responses, and altered cognitive functioning<sup>1</sup>. Recent genome-wide studies in European ancestry populations have revealed both common low risk SNPs and rare high-penetrance copy number variants predisposing to schizophrenia<sup>2-8</sup>. However much of the heritability in 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 genome-wide association study (GWAS) of 3,750 patients with schizophrenia and 6,468 healthy controls genotyped using Affymetrix SNP 6.0 genechips in Han Chinese from three geographic locations (Northern, Central and Southern). In total, 546,561 SNPs were used for statistical analysis after quality control (QC) in 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) (Online Methods).

We carried out a combined analysis of 3 independent genome-wide cohorts as the discovery stage, and the most significant associated SNPs were followed up in additional samples. The initial studies included, i) a genome-wide association study of 1,578 schizophrenia patients

and 1,592 healthy controls from the Northern Han population (Beijing and Shandong Province); ii) a genome-wide analysis of 1,238 cases and 2,856 controls from the Central Han population (Shanghai and Anhui Province); iii) a genome-wide scan of 934 cases and 2,020 controls recruited from the Southern Han population (Guangdong and Guangxi Provinces). Sample descriptions and characteristics can be found in Online Methods and Supplementary Table 1. A meta-analysis (Online Methods) was carried out based on the PCA-adjusted association results of the three cohorts (Supplementary Figure 1). For all the reported SNPs, the p values of the meta-analysis are identical for fixed- and random-effects models of meta-analysis, and no heterogeneity between the groups was found (Supplementary Table 2). Though no SNP showed genome-wide significance ( $P < 5.0 \times 10^{-8}$ ), four SNPs in the chromosome 8p12 (rs1488935,  $P_{\text{gwas-meta}} = 2.81 \times 10^{-6}$ ) and 1q24.2 (rs10489202,  $P_{\text{gwas-meta}} = 2.65 \times 10^{-6}$ ; rs1060041,  $P_{\text{gwas-meta}} = 4.11 \times 10^{-6}$ ; rs11586522,  $P_{\text{gwas-meta}} = 1.53 \times 10^{-6}$ ) regions showed strong association ( $P_{\text{gwas-meta}} < 5.0 \times 10^{-6}$ , Table 1, Supplementary Figure 2, Supplementary Table 2). We followed up the top SNPs in these regions (3 SNPs for 1q24.2 and 2 SNPs for 8p12) in an independent cohort of 4,383 schizophrenia cases and 4,539 healthy controls, which, together with the discovery sample set, we defined as BIOX. We successfully replicated the association findings on 8p12 and 1q24.2, and the combined results of the initial associations and replications showed genome-wide significance (8p12: rs16887244,  $P_{\text{BIOX-meta}} = 1.27 \times 10^{-10}$ , G allele odds ratio (OR)=0.84, 95% confidence interval (CI)=0.79-0.88; rs1488935,  $P_{\text{BIOX-meta}} = 5.06 \times 10^{-9}$ , T allele OR=0.85, 95% CI=0.81-0.90; & 1q24.2: rs10489202,  $P_{\text{BIOX-meta}} = 9.50 \times 10^{-9}$ , 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 (Figure 1).

We further validated our top association signals by obtaining results from a GWAS of the SGENE-plus project (aims to identify genetic variants associating with schizophrenia, and study their impact on phenotype and their interactions with environmental factors contributing to the pathogenesis of the disease); this included 3,830 cases and 14,724 controls of European ancestry. Both SNPs rs16887244 ( $P_{\text{SGENE-plus}} = 0.026$ , G allele OR=0.92, 95% CI=0.85-0.99) and rs1488935 ( $P_{\text{SGENE-plus}} = 0.027$ , T allele OR=0.92, 95% CI=0.85-0.99) showed nominal association with schizophrenia in the European ancestry population, and the direction of the effect was consistent with our findings (Supplementary Table 2 and 3). The effect size of 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 datasets was also carried out (Supplementary Table 4).

On 8p12, both SNPs rs16887244 and rs1488935 demonstrated genome-wide significance in the combined analysis (Table 1, Figure 1a). Controlling for rs16887244, conditional logistic regression analysis revealed that there were no additional association signals (Supplementary Table 5). SNP rs16887244 is located in intron 1 of *LSMI* (MIM: 607281); rs1488935 is located in intron 23 of *WHSC1L1* (MIM: 607083), a gene which is related to the Wolf-Hirschhorn syndrome 1 candidate-like1 gene. We used data from two published expression quantitative trait loci (eQTL) datasets (derived from lymphoblastoid cell lines<sup>9,10</sup> and the brain<sup>11</sup>) to determine whether rs16887244 is associated with the expression of the genes nearby. In the expression data of the lymphoblastoid cell lines (Supplementary Table 6), rs16887244 is nominally associated with the expression of several genes (*ASH2L*, *LSMI*, *BAG4*, *DDHD2*, and *PPAPDC1B*;  $P < 0.05$ ), and in the expression data of the brain (Supplementary Table 7), it is nominally associated with genes *ASH2L*, *DDHD2*, *PPAPDC1B* and *LETM2* ( $P < 0.05$ ).

Notably, rs1488935 is located ~135kb upstream of *FGFR1* (MIM: 136350), the fibroblast growth factor receptor 1 gene. Hippocampal *FGFR1* mRNA expression has been reported to be upregulated in schizophrenia and major depression patients<sup>12</sup> and transgenic mice

expressing a dominant negative mutant [FGFR1(TK-)] from the catecholaminergic, neuron-specific tyrosine hydroxylase (TH) gene promoter have been found to exhibit a schizophrenia-like syndrome<sup>13</sup>. Terwisscha *et al.* in a review found several lines of evidence, including functional plausibility, positional and functional genetic studies, knockout mouse models, effects of using FGF in animals and humans, and associations between FGFs and environmental risk factors of schizophrenia, supporting a role for FGFs in schizophrenia<sup>14</sup>. What is more O'Donovan *et al.* have also reported another fibroblast growth factor receptor gene *FGFR2* associated with schizophrenia ( $P=0.0009$ )<sup>15</sup>.

On 1q24.2, rs10489202 is located in intron 1 of *BRP44* gene, the brain protein 44 gene (Figure 1b), rs1060041 is a coding-synonymous SNP in *DCAF6* (MIM: 610494) and rs11586522 is located in intron 2 region of *GPR161* (MIM: 612250). Controlling for the most significant signal rs10489202, logistic regression analysis indicated that no additional signal was observed in the genomic region around rs10489202 (Supplementary Table 5). In the expression data of the lymphoblastoid cell lines (Supplementary Table 6), rs10489202 is nominally associated with the expression of several genes (*MPZL1*, *DCAF6*, and *TIPRL*;  $P<0.05$ ), however, in the expression data of the brain, we observed no association (Supplementary Table 7). Interestingly, rs10489202 is also ~140kb downstream of *MPZL1* (MIM: 604376), the myelin protein zero like 1 or protein zero-related gene. Tkachev *et al.*<sup>16</sup> found that *MPZL1/PZR* was significantly upregulated in schizophrenia patients compared with healthy controls (1.29-fold,  $P=0.0263$ ). Common SNPs in *MPZL1/PZR* gene were recently reported to be associated with schizophrenia in 523 case-parent triad family of Han Chinese ( $P=0.0017$ )<sup>17</sup>.

Population stratification is a major concern in GWAS that uses case-control design. It has been demonstrated using a set of GWAS datasets that there is a “north-south” population structure in China<sup>18</sup>. The PCA analysis of all our initial study samples also revealed such a structure (Supplementary Figure 3 and 4). We therefore stratified our samples into Northern, Central and Southern groups according to their geographical regions, and carried out GWAS on individual groups, then obtained the combined association using meta-analysis.

The MHC region on chromosome 6p21.3-22.1 has previously been associated with schizophrenia in GWASs of European ancestry populations<sup>4-6</sup>, and these findings have been replicated in a large sample of Han Chinese population<sup>19</sup>. In our initial study, no common SNP in these regions met the selection criteria ( $P_{\text{gwas-meta}} < 5.0 \times 10^{-6}$ ). However, 149 SNPs out of 1,786 (8.34%,  $P_{\text{gwas-meta}} < 0.05$ ) showed nominal association signals with schizophrenia in the MHC region (Supplementary Table 8), and the most significant signal was observed at rs2394514 ( $P_{\text{gwas-meta}} = 1.16 \times 10^{-5}$ ), which is located at 6p22.1. We found no significant association at rs13194053<sup>4,5</sup> and rs6932590<sup>6</sup> from 6p22.1, and rs3131296<sup>6</sup> from 6p21.32, which was previously reported as genome-wide significant associations ( $P < 5.0 \times 10^{-8}$ ) with schizophrenia in European ancestry populations (derived from the NHGRI GWAS Catalog<sup>20</sup>, full SNPs are listed in the Supplementary Table 9). We compared the allele frequencies of these SNPs between HapMap CEU and HCB samples (Supplementary Table 10). The allele frequencies for most of the SNPs obviously differ between the European and Chinese populations. In addition, we measured linkage disequilibrium (LD),  $r^2$  values of these SNPs with the 149 SNPs (in the MHC region with  $P_{\text{gwas-meta}} < 0.05$  in our study) in HapMap CEU and HCB samples (Supplementary Table 11 and 12). SNP rs7749823 which is in LD with rs13194053 ( $r^2=0.644$ ) and rs6932590 ( $r^2=0.623$ ) in the European populations shows association with schizophrenia in our study ( $P_{\text{gwas-meta}} = 4.18 \times 10^{-5}$ , A allele OR=1.64, Risk allele frequency (RAF)=0.97). Interestingly, haplotype analysis of the HapMap CEU data shows the frequency of haplotype TTA of rs13194053-rs6932590-rs7749823 is 0.808. As both T alleles of rs13194053 and rs6932590 are risk alleles for schizophrenia in the European population studies, our association signal

with the A allele of rs7749823 of the three marker haplotype in the Han Chinese population is consistent with the European findings.

We also examined other known schizophrenia susceptibility loci (*NRGN*, 11q24.2; and *TCF4*, 18q21.2) which were reported as genome-wide significant in the published studies of schizophrenia (Supplementary Table 9). However, none of the SNPs within the loci in our dataset show association with schizophrenia. However, genetic heterogeneity obviously exists for schizophrenia risk variants across ethnicities (Supplementary Table 10). This reinforces the need for genome-wide association studies of schizophrenia to be carried out in diverse populations.

In summary we have identified common genetic variants in 8p12 and 1q24.2 to be associated with schizophrenia in the Han Chinese population. Several promising candidate genes are implicated in these two regions. This makes it difficult to decide precisely which genes contain the causative variants in these two chromosomal regions. Nevertheless identification of these novel common genetic risk variants that predispose to schizophrenia is an encouraging first step in a process that has the potential to be translated into improved methods for the prediction/treatment of the disease.

## Online Methods

### Subjects

All schizophrenia patients analyzed in the BIOX study were interviewed by 2 independent psychiatrists, diagnosed according to DSM-IV criteria and had a 2-year history of schizophrenia. All cases met 2 criteria: preoccupation with one or more delusions and frequent auditory hallucinations; however, none of the following symptoms were prominent: disorganized speech, disorganized or catatonic behavior, or flat or inappropriate affect. 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 Chinese Han volunteers who were requested to reply to a written invitation to evaluate their medical history. Practice lists of controls were screened for potentially suitable volunteers by excluding subjects with major mental illness.

The discovery phase (BIOX GWAS) includes three datasets: the Northern Han sample set of 1,578 cases and 1,592 controls were recruited from Beijing and Shandong Province; the Central Han sample set of 1,238 cases and 2,856 controls were recruited from Shanghai and Anhui Province; and the Southern Han sample set of 934 cases and 2,020 controls were recruited from Guangdong and Guangxi Provinces. Sample descriptions can be found in the Supplementary Table 1.

The replication stage (BIOX FOLLOW-UP) included 4,383 schizophrenia cases and 4,539 healthy controls recruited from Shanghai. All subjects in the replication stage were unrelated and born in Shanghai and their parents had to be local residents of Shanghai as well. This replication sample is 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 the Supplementary Table 1.

The SGENE-plus samples included 513 patients and 471 controls from Denmark/Copenhagen; 93 schizophrenia patients and 88 unrelated controls from England; 182 schizophrenia patients and 197 controls from Finland; 1048 patients and 971 controls from Germany; 531 cases and 11,615 controls from Iceland; 84 patients and 89 controls from



Italy; 693 patients and 629 controls from The Netherlands; 658 schizophrenia cases and 661 controls from Scotland. All case samples were diagnosed according to ICD-10 or DSM-IV criteria. These samples are the same as the genome-wide typed samples used in a previous Nature publication<sup>6</sup> with two exceptions: slight changes in the Icelandic schizophrenia patients used (~10% difference), and two newly genome-wide typed groups from 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 were listed in the Supplementary Sample Information.

### DNA extraction

EDTA anti-coagulated venous blood samples were collected from all participants. Genomic DNA was extracted from peripheral blood lymphocytes by standard procedures using Flexi Gene DNA kits (Fuji), diluted to working concentrations of 50 ng/μl for SNP chip genotyping, and 10~20 ng/μl for replication genotyping.

### GWAS genotyping and quality control

The genome-wide scan 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 a Contrast QC < 0.4 from further data analysis. Genotype data were generated using the birdseed algorithm<sup>21</sup>. 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 geographical group. SNPs with minor allele frequency (MAF) < 1% or significant deviation from Hardy Weinberg Equilibrium (HWE;  $P < 1 \times 10^{-3}$ ) in controls were also excluded. SNPs passing QC in the Northern, Central and Southern Han groups were used for further analysis. After QC, there were 546,561 SNPs remaining for the combined initial study.

### SNP selection criteria and replication genotyping

We selected all SNPs with  $P_{\text{initial-meta}} < 5.0 \times 10^{-6}$  for the replication study, according to a previous GWAS<sup>22</sup>. To ensure that there were at least 2 SNPs included for each region, we added the 2<sup>nd</sup> most significant SNP among the 8p12 findings for replication. Genotyping for the replication study was performed using the Ligation Detection Reaction (LDR) method<sup>23,24</sup>, with technical support from Shanghai Biowing Applied Biotechnology Company.

### Analysis of population substructure

Population substructure was evaluated using principal components analysis as implemented in the software EIGENSTRAT<sup>25,26</sup>. 20 principal components (PCs), some of which will reflect ancestry differences among subjects, were generated for each subject. Logistic regression was used to determine whether there was a significant difference in PC scores between cases and controls; significant PCs were used as covariates in the association analysis to correct for population stratification. We identified 6 significant PCs (for the Northern data set), 8 significant PCs (for the Central data set), and 4 significant PCs (for the Southern data set).

### Imputation analysis of untyped SNPs

We selected the two new identified potential susceptibility loci (1q24.2, rs10489202 ± 350kb and 8p12, rs16887244 ± 350kb) for imputation. Hapmap SNPs in the 2 regions were imputed using MACH 1.0. Phased haplotypes for 90 CHB+JPT subjects (180 haplotypes)

were used as the reference for imputing genotypes. Any SNP imputed with information content  $r^2 < 0.3$  was excluded from association analysis because of lack of power. The criteria for SNP QC filtering were the same as the genotyped ones.

## Statistical methods

Association of single SNPs with schizophrenia was tested by logistic regression using PLINK<sup>27</sup>, separately for the Northern, Central and Southern data sets, correcting for principal component scores that had statistically significant differences between cases and controls. The Hardy-Weinberg-Equilibrium analysis was performed using PLINK, and Haploview<sup>28</sup> was used for the genome wide P value plot. The Q-Q plot was created using the R package. The regional plots were generated using LocusZoom (see URLs). In the replication study, allelic association analysis was conducted using SHEsis<sup>29</sup>. A meta-analysis using the random-effect model was carried out based on the PCA-adjusted association results of the three cohorts in the initial GWAS, the inflation factor value was 0.97. The meta-analysis was conducted using PLINK. Heterogeneity across the data sets was evaluated using Cochran's Q test. The meta-analysis was carried out using the Mantel-Haenzel method with a random-effects model<sup>30</sup>. Conditional logistic regression was used to test for independent effects of an individual SNP<sup>22,31</sup>. Based on the sample size of the BIOX and SGENE-plus datasets, we compared the statistical power between BIOX and SGENE-plus (Supplementary Table 4;  $\alpha = 5 \times 10^{-8}$ )<sup>5,32</sup>.

## Expression quantitative trait loci (eQTL) analysis

Expression profiles were analyzed within two eQTL datasets (lymphoblastoid cell lines and brain). We downloaded the expression data sets from NCBI GEO (<http://www.ncbi.nlm.nih.gov/geo/>). The first dataset, published by Stranger and colleagues<sup>9,10</sup>, consists of gene expression profiles generated using RNA extracted from lymphoblastoid cell lines of all 210 unrelated HapMap individuals from four populations (CEU: 60 Utah residents with ancestry from northern and western Europe; CHB: 45 Han Chinese in Beijing; JPT: 45 Japanese in Tokyo; YRI: 60 Yoruba in Ibadan, Nigeria). Expression analysis was performed using Sentrix Human-6 Expression BeadChip (Illumina, San Diego, California, United States). The SNP genotypes from phase II HapMap were used in the analysis. The second eQTL dataset<sup>11</sup> consists of gene expression profiles for frozen tissue samples obtained from four brain regions (cerebellum, pons, 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 performed using Infinium HumanHap550 BeadChips. The expression data were normalized and log transformed as described in the original reports<sup>9-11</sup>. 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 lines dataset, the analyses were conducted for each population and the combined dataset. For the brain dataset, the analysis of each tissue region was performed separately.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We are deeply grateful to all the participants as well as to the doctors working on this project. The authors also thank editors and anonymous reviewers for their valuable comments on the manuscript. This work was supported by the 863 Program (2006AA02A407, 2009AA022701), the Natural Science Foundation of China (31000553), the 973 Program (2010CB529600), the Funding Awarded to Authors for National Excellence of Doctoral Dissertation Research (201026), the Program for New Century Excellent Talents in University (NCET-09-0550), the Shanghai

Municipal Health Bureau program (2008095), the Shanghai Changning Health Bureau program (2008406002), the Shanghai Municipal Commission of Science and Technology Program (09DJ1400601), and National Key Technology R&D Program (2006BAI05A09), and the Shanghai Leading Academic Discipline Project (B205).

## References

1. Craddock N, O'donovan MC, Owen MJ. The genetics of schizophrenia and bipolar disorder: dissecting psychosis. *Journal of Medical Genetics*. 2005; 42:193–204. [PubMed: 15744031]
2. McCarthy SE, et al. Microduplications of 16p11. 2 are associated with schizophrenia. *Nature genetics*. 2009; 41:1223–1227. [PubMed: 19855392]
3. O'Donovan MC, et al. Identification of loci associated with schizophrenia by genome-wide association and follow-up. *Nature genetics*. 2008; 40:1053–1055. [PubMed: 18677311]
4. Purcell SM, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*. 2009; 460:748–752. [PubMed: 19571811]
5. Shi J, et al. Common variants on chromosome 6p22. 1 are associated with schizophrenia. *Nature*. 2009; 460:753–757. [PubMed: 19571809]
6. Stefansson H, et al. Common variants conferring risk of schizophrenia. *Nature*. 2009; 460:744–747. [PubMed: 19571808]
7. Vacic V, et al. Duplications of the neuropeptide receptor gene VIPR2 confer significant risk for schizophrenia. *Nature*. 2011; 471:499–503. [PubMed: 21346763]
8. Xu B, et al. Strong association of de novo copy number mutations with sporadic schizophrenia. *Nature genetics*. 2008; 40:880–885. [PubMed: 18511947]
9. Stranger BE, et al. Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science*. 2007; 315:848–853. [PubMed: 17289997]
10. Stranger BE, et al. Population genomics of human gene expression. *Nature genetics*. 2007; 39:1217–1224. [PubMed: 17873874]
11. Gibbs JR, et al. Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. *PLoS Genet*. 2010; 6:e1000952. [PubMed: 20485568]
12. Gaughran F, Payne J, Sedgwick PM, Cotter D, Berry M. Hippocampal FGF-2 and FGFR1 mRNA expression in major depression, schizophrenia and bipolar disorder. *Brain research bulletin*. 2006; 70:221–227. [PubMed: 16861106]
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. *Journal of neurochemistry*. 2006; 97:1243–1258. [PubMed: 16524369]
14. Terwisscha SAF, Bakker SC, Kahn RS. Fibroblast Growth Factors in Schizophrenia. *Schizophrenia bulletin*. 2010; 36:1157–1166. [PubMed: 19429845]
15. O'Donovan MC, et al. Analysis of 10 independent samples provides evidence for association between schizophrenia and a SNP flanking fibroblast growth factor receptor 2. *Molecular psychiatry*. 2008; 14:30–36. [PubMed: 18813210]
16. Tkachev D, et al. Oligodendrocyte dysfunction in schizophrenia and bipolar disorder. *The Lancet*. 2003; 362:798–805.
17. He G, et al. MPZL1/PZR, a novel candidate predisposing schizophrenia in Han Chinese. *Molecular psychiatry*. 2006; 11:748–751. [PubMed: 16702974]
18. Chen J, et al. Genetic structure of the Han Chinese population revealed by genome-wide SNP variation. *The American Journal of Human Genetics*. 2009; 85:775–785.
19. Li T, et al. Common Variants in Major Histocompatibility Complex Region and TCF4 Gene Are Significantly Associated with Schizophrenia in Han Chinese. *Biological psychiatry*. 2010; 68:671–673. [PubMed: 20673877]
20. Hindorf LA, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proceedings of the National Academy of Sciences*. 2009; 106:9362–9367.
21. Korn JM, et al. Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. *Nature genetics*. 2008; 40:1253–1260. [PubMed: 18776909]



22. Chen ZJ, et al. Genome-wide association study identifies susceptibility loci for polycystic ovary syndrome on chromosome 2p16. 3, 2p21 and 9q33. *Nature Genetics*. 2011; 43:55–59. [PubMed: 21151128]
23. Thomas G, et al. Capillary and microelectrophoretic separations of ligase detection reaction products produced from low - abundant point mutations in genomic DNA. *Electrophoresis*. 2004; 25:1668–1677. [PubMed: 15188256]
24. Yi P, et al. PCR/LDR/capillary electrophoresis for detection of single - nucleotide differences between fetal and maternal DNA in maternal plasma. *Prenatal diagnosis*. 2009; 29:217–222. [PubMed: 19177453]
25. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genet*. 2006; 2:e190. [PubMed: 17194218]
26. Price AL, et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics*. 2006; 38:904–909. [PubMed: 16862161]
27. Purcell S, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*. 2007; 81:559–575.
28. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*. 2005; 21:263–265. [PubMed: 15297300]
29. Yong Y, Lin HE. SHEsis, a powerful software platform for analyses of linkage disequilibrium, haplotype construction, and genetic association at polymorphism loci. *Cell research*. 2005; 15:97–98. [PubMed: 15740637]
30. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies. *J Natl Cancer Inst*. 1959; 22:719–748. [PubMed: 13655060]
31. Petukhova L, et al. Genome-wide association study in alopecia areata implicates both innate and adaptive immunity. *Nature*. 2010; 466:113–117. [PubMed: 20596022]
32. Menashe I, Rosenberg PS, Chen BE. PGA: power calculator for case-control genetic association analyses. *BMC genetics*. 2008; 9:36. [PubMed: 18477402]

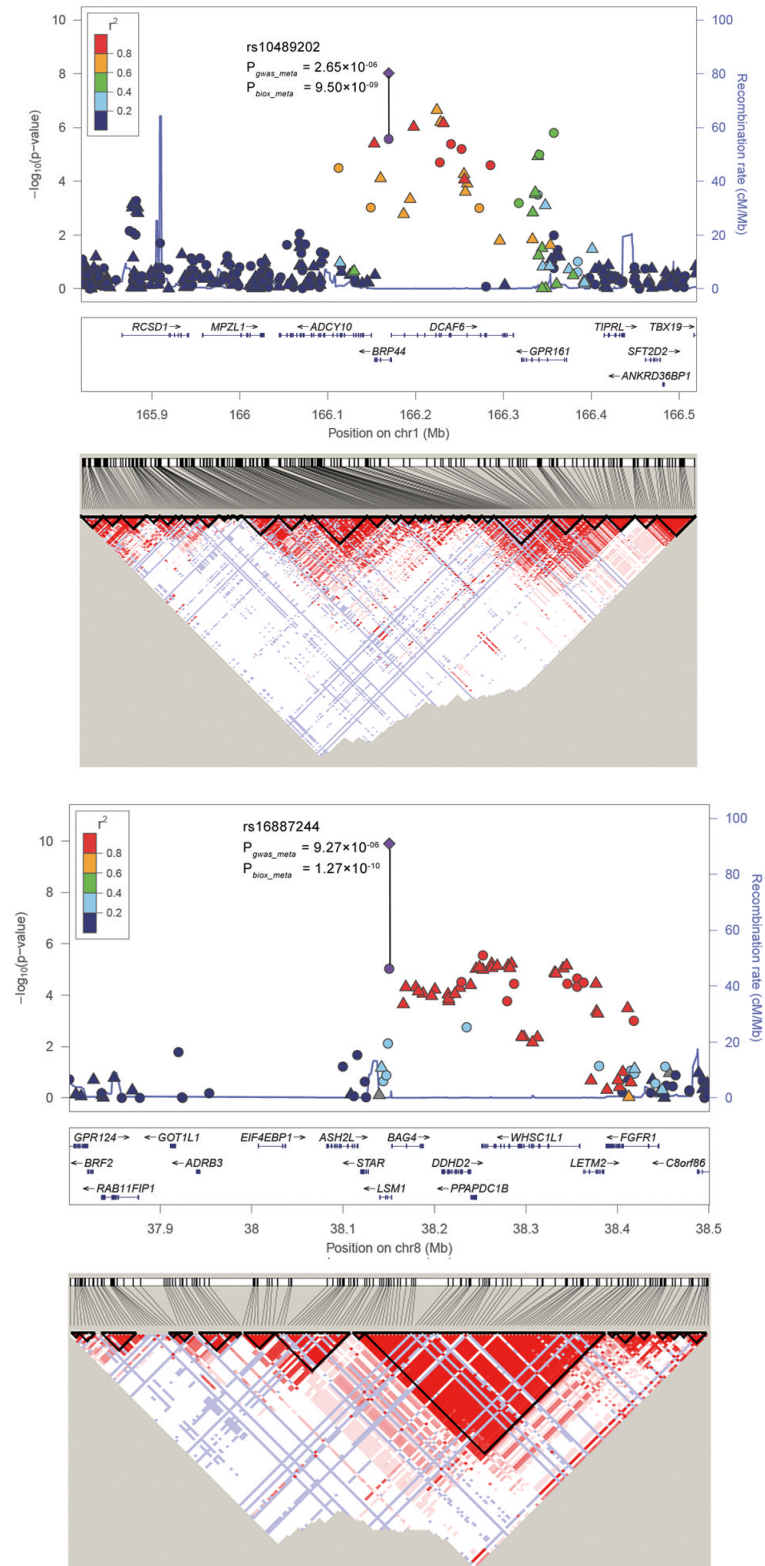


Figure 1.

Regional plots of the 2 schizophrenia loci (8p12 and 1q24.2). Results ( $-\log_{10} P$ ) are shown for SNPs for the region flanking 350 kb on either side of the marker SNPs.  $P_{\text{gwas\_meta}}$  is for the combined results of the three initial datasets, represented by the circle (for genotyped SNPs) and the triangle (for imputed SNPs), and  $P_{\text{bioX\_meta}}$  is for the combined results of the initial and follow-up dataset (BIOX), represented by the diamond. The marker SNP is shown in purple and the  $r^2$  values of the rest of the SNPs are indicated by different colors. The genes within the relevant region are annotated and shown as arrows. The D LD block structure of the related surrounding region according to the CHB+JPT HapMap data are provided at the bottom.

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

CHR	SNP	BP	Frequency <sup>a</sup>	BIOX GWAS									
				NORTHERN HAN (1578 cases, 1592 controls)	CENTRAL HAN (1238 cases, 2856 controls)	SOUTHERN HAN (934 cases, 2020 controls)	INITIAL META	BIOX FOLLOW-UP (4383 cases, 4539 controls)		BIOX GWAS +FOLLOW-UP			
	Allele <sup>b</sup>	OR (95% CI)	P*	OR (95% CI)	P*	OR (95% CI)	P*	OR (95% CI)	Allele	OR (95% CI)	P <sub>rep</sub>	P <sub>BIOX-meta</sub>	
1	rs10489202	166169703	0.141	A (1.084-1.494)	0.003	1.309 (1.122-1.528)	6.19×10 <sup>-4</sup>	1.180 (0.953-1.461)	0.130	2.65×10 <sup>-6</sup>	A (1.094-1.293)	4.76×10 <sup>-5</sup>	9.50×10 <sup>-9</sup>
1	rs1060041	166240600	0.139	T (1.083-1.493)	0.003	1.324 (1.133-1.547)	4.06×10 <sup>-4</sup>	1.136 (0.916-1.410)	0.245	4.11×10 <sup>-6</sup>	T (1.025-1.211)	0.012	5.31×10 <sup>-7</sup>
1	rs11586522	166357655	0.160	T (1.099-1.483)	0.001	1.267 (1.101-1.459)	9.77×10 <sup>-4</sup>	1.181 (0.956-1.46)	0.123	1.53×10 <sup>-6</sup>	T (0.957-1.129)	0.362	1.17×10 <sup>-4</sup>
8	rs16887244	38150502	0.317	G (0.747-0.972)	0.017	0.809 (0.718-0.912)	5.16×10 <sup>-4</sup>	0.876 (0.745-1.029)	0.107	9.27×10 <sup>-6</sup>	G (0.781-0.888)	2.23×10 <sup>-8</sup>	1.27×10 <sup>-10</sup>
8	rs1488935	38252950	0.316	T (0.737-0.962)	0.011	0.795 (0.703-0.898)	2.30×10 <sup>-4</sup>	0.871 (0.742-1.023)	0.093	2.81×10 <sup>-6</sup>	T (0.816-0.929)	3.42×10 <sup>-5</sup>	5.06×10 <sup>-9</sup>

<sup>a</sup>Frequency is the average minor allelic frequency of controls in BIOX data

<sup>b</sup>minor allele

\* These P values were adjusted by PCA methods in each GWAS of geographical sample groups.