

Genome-wide association analysis identifies 30 new susceptibility loci for schizophrenia

Zhiqiang Li^{1–4,22} , Jianhua Chen^{2,5,22}, Hao Yu^{6–8}, Lin He^{2,4}, Yifeng Xu⁵, Dai Zhang^{6,7,9}, Qizhong Yi¹⁰, Changgui Li¹¹, Xingwang Li², Jiawei Shen², Zhijian Song², Weidong Ji^{4,12}, Meng Wang², Juan Zhou², Boyu Chen², Yahui Liu², Jiqiang Wang², Peng Wang¹³, Ping Yang¹³, Qingzhong Wang², Guoyin Feng⁵, Benxiu Liu¹⁴, Wensheng Sun¹⁴, Baojie Li², Guang He², Weidong Li², Chunling Wan², Qi Xu¹⁵, Wenjin Li², Zujia Wen², Ke Liu², Fang Huang², Jue Ji², Stephan Ripke^{16–18}, Weihua Yue^{6,7}, Patrick F Sullivan^{19,20} , Michael C O'Donovan²¹  & Yongyong Shi^{1–4,10,12} 

We conducted a genome-wide association study (GWAS) with replication in 36,180 Chinese individuals and performed further transancestry meta-analyses with data from the Psychiatry Genomics Consortium (PGC2). Approximately 95% of the genome-wide significant (GWS) index alleles (or their proxies) from the PGC2 study were overrepresented in Chinese schizophrenia cases, including ~50% that achieved nominal significance and ~75% that continued to be GWS in the transancestry analysis. The Chinese-only analysis identified seven GWS loci; three of these also were GWS in the transancestry analyses, which identified 109 GWS loci, thus yielding a total of 113 GWS loci (30 novel) in at least one of these analyses. We observed improvements in the fine-mapping resolution at many susceptibility loci. Our results provide several lines of evidence supporting candidate genes at many loci and highlight some pathways for further research. Together, our findings provide novel insight into the genetic architecture and biological etiology of schizophrenia.

Schizophrenia (MIM181500) is a chronic, severe and disabling brain disorder that affects approximately 1% of the worldwide population and imposes an enormous burden on society^{1,2}. It is a highly heritable psychiatric disorder (with an estimated heritability of 70–85%)³ with a substantial polygenic component including thousands of common alleles with small effects that contribute to disease risk⁴. Approximately 33–50% of the genetic risk of schizophrenia has been captured by common alleles in GWAS⁵. The evidence to date suggests that many risk alleles for common schizophrenia-associated genetic loci may be shared across ancestry groups, but others may be population specific because of differing causal variants or linkage disequilibrium (LD) patterns in populations of different ancestries⁶. Previous GWAS have identified more than 110 schizophrenia-associated loci and have substantially advanced understanding of

this condition^{5,7–13}. In particular, the most recent and largest schizophrenia GWAS (from the Schizophrenia Working group of the Psychiatric Genomics Consortium, PGC2) which, with discovery and extension, included a total of 36,989 schizophrenia cases and 113,075 controls, has identified 128 independent genome-wide significant associations spanning 108 loci⁷.

However, a large proportion of the genetic factors underlying schizophrenia remain unknown. Most of the heritability of schizophrenia is not yet attributable to specific loci; only 3.5% of the liability can be explained by GWS loci⁷. Moreover, to date, most GWAS participants with schizophrenia are of European descent. Thus, although the PGC2 report includes samples from East Asia, the proportions are small (approximately 5% and 3% for cases and controls, respectively). Large-scale GWAS including individuals of non-European descent

¹Affiliated Hospital of Qingdao University and Biomedical Sciences Institute of Qingdao University (Qingdao Branch of SJTU Bio-X Institutes), Qingdao University, Qingdao, China. ²Bio-X Institutes, Key Laboratory for the Genetics of Developmental and Neuropsychiatric Disorders (Ministry of Education), Collaborative Innovation Center for Brain Science, Shanghai Jiao Tong University, Shanghai, China. ³Institute of Social Cognitive and Behavioral Sciences, Shanghai Jiao Tong University, Shanghai, China. ⁴Institute of Neuropsychiatric Science and Systems Biological Medicine, Shanghai Jiao Tong University, Shanghai, China. ⁵Shanghai Key Laboratory of Psychotic Disorders, Shanghai Mental Health Center, Shanghai Jiao Tong University School of Medicine, Shanghai, China. ⁶Tsinghua-Peking Joint Center for Life Sciences, School of Life Sciences, Tsinghua University, Beijing, China. ⁷Institute of Mental Health, Sixth Hospital, Peking University, Beijing, China. ⁸Department of Psychiatry, Jining Medical University, Jining, China. ⁹Peking-Tsinghua Joint Center for Life Sciences/PKU-IDG/McGovern Institute for Brain Research, Peking University, Beijing, China. ¹⁰Department of Psychiatry, First Teaching Hospital of Xinjiang Medical University, Urumqi, China. ¹¹Shandong Provincial Key Laboratory of Metabolic Disease and Metabolic Disease Institute of Qingdao University, Qingdao, China. ¹²Changning Mental Health Center, Shanghai, China. ¹³Wuhu Fourth People's Hospital, Wuhu, China. ¹⁴Longquan Mountain Hospital of Guangxi Province, Liuzhou, China. ¹⁵National Laboratory of Medical Molecular Biology, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China. ¹⁶Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts, USA. ¹⁷Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA. ¹⁸Department of Psychiatry and Psychotherapy, Charité-Universitätsmedizin, Berlin, Germany. ¹⁹Departments of Genetics and Psychiatry, University of North Carolina, Chapel Hill, North Carolina, USA. ²⁰Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. ²¹MRC Centre for Neuropsychiatric Genetics and Genomics, Division of Psychological Medicine and Clinical Neurosciences, School of Medicine, Cardiff University, Cardiff, UK. ²²These authors contributed equally to this work. Correspondence should be addressed to Y.S. (shiyongyong@gmail.com).

Received 17 April; accepted 19 September; published online 9 October 2017; doi:10.1038/ng.3973

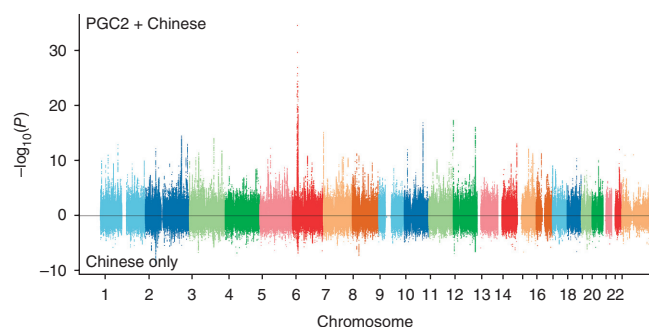


Figure 1 Comparison of Manhattan plots for the Chinese and transancestry analyses. Manhattan plot of results from the Chinese-only (7,699 schizophrenia cases and 18,327 controls) and PGC2-plus-Chinese transancestry (43,175 cases and 65,166 controls) analyses. $-\log_{10}P$ values for PGC2 plus Chinese transancestry analyses and $\log_{10}P$ values for Chinese-only analyses are shown.

are essential for extending understanding of the genetic architecture of schizophrenia in the human population as a whole, for testing the generalizability of the results from European populations regarding this global disorder⁶ and for identifying population-specific risk factors, should they exist.

To identify additional schizophrenia susceptibility loci and to gain a better understanding of the genes and biological pathways implicated in schizophrenia, we performed a GWAS including 7,699 cases and 18,327 controls of Chinese ancestry, as well as a transancestry GWAS meta-analysis with PGC2 (43,175 cases and 65,166 controls in total). The candidate loci found in each analysis were then studied in an independent replication sample of 4,384 schizophrenia cases and 5,770 controls of Chinese ancestry.

RESULTS

Results of GWAS screening in the Chinese population

We first conducted a GWAS for schizophrenia in the Chinese population (in comparison to the discovery phase of our prior GWAS report¹⁰, the number of cases was doubled, and the number of controls was tripled). After systematic quality control (QC) analysis and imputation to the 1000 Genomes Project data (Online Methods), we assessed the associations of 5,107,227 genetic variants in 7,699 schizophrenia cases and 18,327 controls (**Supplementary Table 1**). The primary GWAS comprised three samples that were genotyped on different platforms: 4,175 cases and 10,470 controls genotyped with the Affymetrix Genome-Wide Human SNP Array 6.0 (SNP6.0); 2,472 cases and 5,928 controls genotyped with the Affymetrix Axiom Genome-Wide CHB1 Array Plate (CHB1); and 1,052 cases and 1,929 controls genotyped with CHB1 or the Illumina 1M Array (1M). Principal component analysis (PCA) was used to assess population substructure (**Supplementary Fig. 1** and Online Methods). For each subset, association testing was conducted with logistic regression including ancestry principal components (PCs) as covariates to adjust for population stratification. The results were combined with inverse-variance-weighted meta-analysis (based on a fixed-effects model). The genomic inflation factor (λ_{GC}) was 1.22, and the $\lambda_{1,000}$ (a scaled value to 1,000 cases and 1,000 controls) was 1.02. We conducted LD-score regression analysis¹⁴ to distinguish the relative contributions of confounding bias and polygenicity. The LD-score regression intercept was 1.07 (s.e. = 0.01), and the slope was greater than zero, thus suggesting that most of the increase in the mean χ^2 statistic was from polygenic architecture rather than population stratification, in agreement with the previously documented polygenic nature of schizophrenia inheritance^{7,15}. However, given this modest

elevation in the intercept, we further corrected the meta-analysis statistics for residual test-statistic inflation^{14,16} (Online Methods). Quantile–quantile and Manhattan plots are shown in **Supplementary Figures 2 and 3**. In this analysis (**Fig. 1**), we observed 66 GWS variants in a region previously reported to be associated with schizophrenia (2p16.1)^{11,12}.

The proportion of variance in susceptibility to schizophrenia explained by genome-wide SNP genotypes for individuals of Han Chinese ancestry (Online Methods) was estimated to be 31.5% (s.e. = 1.9%), assuming a population risk of 0.01. This result was similar to the corresponding estimate for European samples (33%)⁵, thus providing further evidence of the highly polygenic nature of schizophrenia beyond that in previous studies^{7,15}.

Results of the Chinese and PGC2 genome-wide meta-analysis

We performed a meta-analysis of the Chinese GWAS samples (7,699 schizophrenia cases and 18,327 controls) (denoted Chinese GWAS) and PGC2 GWAS samples (35,476 schizophrenia cases and 46,839 controls) to explore the effects of power and heterogeneity. A total of 4,303,606 genetic variants were common to the two data sets and were retained in the combined analysis. For combining the data, we used a fixed-effects model, but for variants with pronounced heterogeneity ($I^2 > 75\%$)¹⁷, we used a random-effects model to allow for the possibility that the presence of heterogeneity might result in test-statistic inflation. In our final result, the λ_{GC} was 1.50, and the $\lambda_{1,000}$ was less than 1.01. The deviation of the observed statistics from the null hypothesis was less than that expected under a polygenic model for schizophrenia^{7,18}. Quantile–quantile and Manhattan plots are shown in **Supplementary Figures 4 and 5**. In the combined analysis, we detected 5,618 SNPs surpassing the threshold for GWS for association with schizophrenia. These SNPs mapped to 104 physically distinct associated regions, as defined by clumping the variants by using $r^2 > 0.1$ and merging the LD-independent variants within 250 kb (**Fig. 1** and **Supplementary Table 2**).

Results of the combined analysis with replication samples

We then obtained association results from an independent Chinese cohort of 4,384 schizophrenia cases and 5,770 controls¹⁹ (**Supplementary Table 1**) for LD-independent SNPs with $P < 1 \times 10^{-5}$ in the Chinese-only GWAS meta-analysis or with $P < 5 \times 10^{-7}$ in the Chinese and PGC2 GWAS meta-analysis (Online Methods).

The combined analysis of the Chinese GWAS and replication samples resulted in a data set of 12,083 cases and 24,097 controls. Seven loci were GWS for association with schizophrenia in the meta-analysis of individuals of Chinese ancestry. Of those loci, three have been previously reported to be associated with schizophrenia (**Supplementary Fig. 6**), and the other four are novel: rs2073499 at 3p21.31 (odds ratio (OR) = 0.899, fixed-effects meta-analysis $P = 2.61 \times 10^{-8}$), rs7757969 at 6q21 (OR = 1.110, $P = 4.82 \times 10^{-8}$), rs4479915 at 6q27 (OR = 0.876, $P = 4.82 \times 10^{-9}$) and rs11534004 at 7q31.1 (OR = 0.890, $P = 1.71 \times 10^{-8}$) (**Fig. 2**). Four additional loci were significant at $P < 1 \times 10^{-5}$ in the Chinese GWAS meta-analysis and showed nominal evidence of replication ($P < 0.05$) but were not GWS in the combined analysis. Results for all tested SNPs are presented in **Supplementary Table 3**.

The combined results of the transancestry meta-analysis (43,175 schizophrenia cases and 65,166 controls) and replication samples (4,384 schizophrenia cases and 5,770 controls) identified a total of 109 GWS loci (**Supplementary Table 4** and **Supplementary Data 1**). Of the 109 loci, 83 had previously been reported, and 26 loci were novel.

Together, the above results identified 124 SNPs that were GWS and were associated with schizophrenia. The SNPs mapped to 113

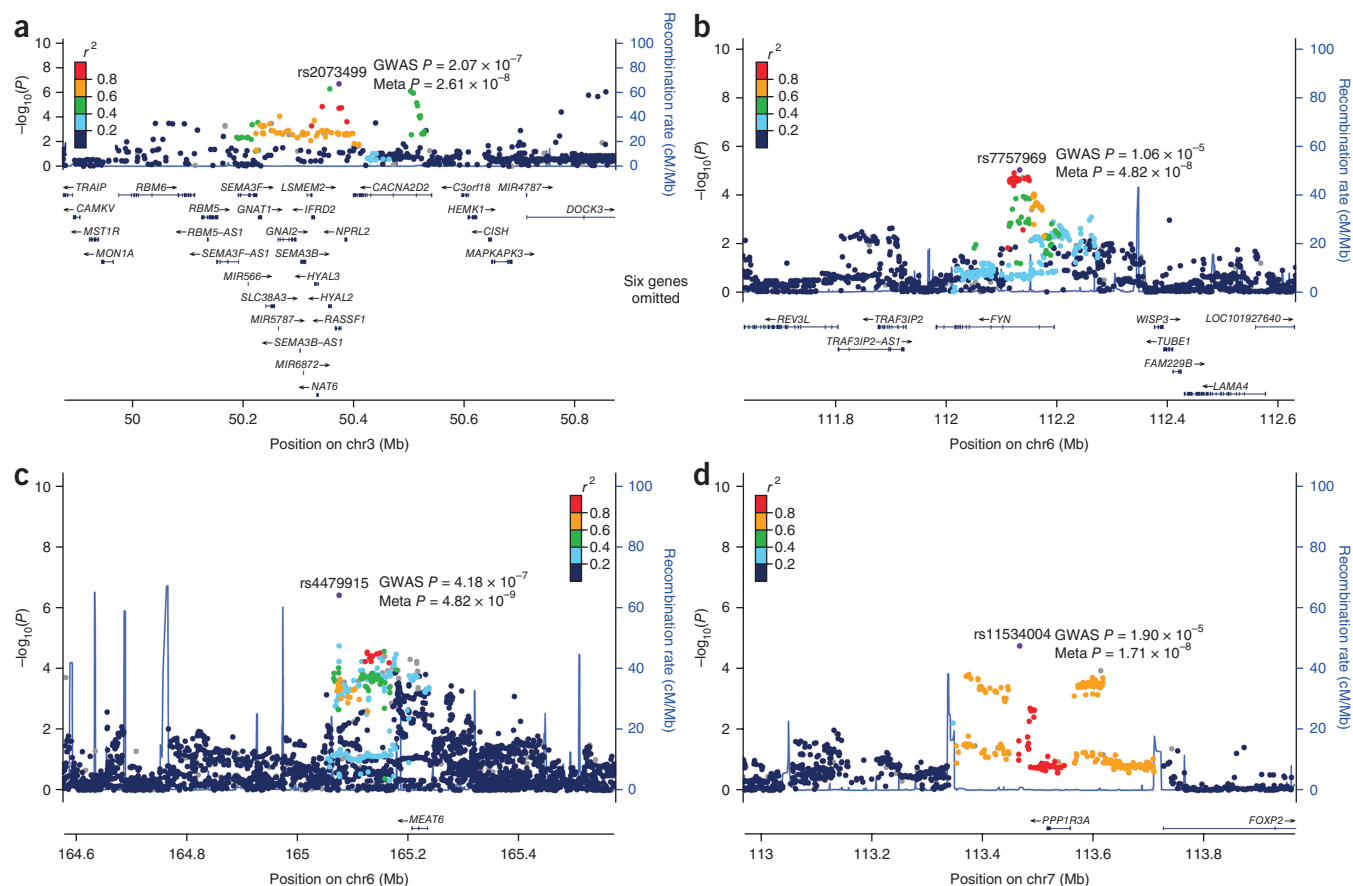


Figure 2 Regional plots for novel GWS loci in Chinese people. (a) rs2073499 at 3p21.31. (b) rs7757969 at 6q21. (c) rs4479915 at 6q27. (d) rs11534004 at 7q31.1. Meta, meta-analysis; chr, chromosome. $-\log_{10}P$ values are shown for SNPs for the region 500 kb on either side of the marker SNPs. The index SNP is shown in purple, and the r^2 values of the other SNPs are indicated by color. The r^2 values were established on the basis of 1000 Genomes data (November 2014). P values for the GWAS stage are shown with circles, and P values for the meta-analysis combining all data sets are shown with text. The genes within the relevant regions are annotated and shown as arrows.

physically distinct loci: four loci were GWS only in the Chinese-only analysis, 106 were GWS only in the transancestry analysis, and three were present in both analyses (Supplementary Table 5). Of the 113 associated loci, 30 have not been previously reported (Table 1), four of which were GWS in the Chinese sample but not the transancestry analysis. In addition, at three of the previously reported loci, the GWS SNPs in the present study were in low LD with the previously identified GWS SNPs ($r^2 < 0.1$ in both the European and Chinese populations), thus possibly suggesting independent signals in these regions (Supplementary Table 5).

Similarities and differences across ancestries

Of the 108 loci (128 index SNPs) identified in the PGC2 report⁷, we were able to investigate 103 loci (117 index SNPs or their proxies) that were in common between PGC2 and Chinese data sets (Supplementary Table 6). Of these, the PGC2-associated risk alleles were overrepresented in Chinese cases at 109 SNPs (from 98 loci), and at 58 SNPs (from 56 loci) this overrepresentation achieved nominal significance ($P < 0.05$). In transancestry meta-analyses, 85 SNPs at 78 loci continued to be GWS. It is known that the random-effects model might be overly conservative²⁰, and therefore on an exploratory basis, we performed a fixed-effects model meta-analysis for all these SNPs regardless of the existence of heterogeneity. Under the fixed-effects model, an additional eight SNPs (93 in total) at eight loci (86 loci in

total) were GWS in the combined analysis. However, the results for the GWS SNPs indicated by fixed-effects meta-analysis and with evidence of heterogeneity should be interpreted with caution. Nevertheless, this finding suggested that the schizophrenia susceptibility loci identified in European samples were applicable to the Chinese sample. Moreover, the transancestry meta-analyses also confirmed two GWS loci (8p12 and 7q11.22) identified in our previous reports^{10,21}.

Regarding the seven GWS index SNPs analyzed in the Chinese-only analysis in this study, three replicated at $P < 0.05$ in the PGC2 data set but showed significant heterogeneity (Higgins and Thompson I^2 index $> 75\%$) across populations and were not GWS in transancestry meta-analyses. In addition, one of the index SNPs (rs78681500) was absent in the PGC2 data set, owing to its rarity (minor allele frequency (MAF) $< 1\%$). Of the 117 GWS index SNPs identified in the transancestry analysis, all showed the same direction of effect across ancestries, and the I^2 was less than 75%.

We next assessed the genome-wide congruence of risk alleles across the PGC2 and Chinese GWAS data sets for LD-clumped independent SNPs²². For the schizophrenia-associated SNPs ($P \leq 0.0001$) identified in the Chinese data set, we observed a highly significant excess of directional concordance in the PGC2 data set (67.7%, binomial test $P = 3.06 \times 10^{-7}$). For the SNPs demonstrating weaker evidence of an association with schizophrenia ($0.0001 < P \leq 0.05$), we also observed an excess of consistency in the direction of effect. In contrast, for the

Table 1 Novel schizophrenia GWS loci and notable genes

Chromosome	SNP	Position	P value	Notable gene(s) ^a
2	rs999494	73157395	2.40×10^{-10}	<i>EMX1</i> (N, D)
2	rs62152284	104984387	5.86×10^{-9}	<i>LOC100287010</i> (N)
2	rs6430491	134840967	9.55×10^{-10}	<i>MIR3679</i> (N)
3	rs10510653	32058559	2.54×10^{-8}	<i>GPD1L</i> (Q), <i>ZNF860</i> (N)
3	rs2073499	50374293	2.61×10^{-8}	<i>HYAL3</i> (Q), <i>RASSF1</i> (N)
4	rs11722779	103827488	3.40×10^{-8}	<i>BDH2</i> (Q), <i>CENPE</i> (Q), <i>CISD2</i> (Q), <i>KRT8P46</i> (Q), <i>LRR37A15P</i> (Q), <i>NHEDC1</i> (N), <i>SLC9B1</i> (Q)
5	rs10940346	49806042	1.11×10^{-8}	<i>EMB</i> (N, Q)
5	rs2247870	90151589	2.54×10^{-8}	<i>ADGRV1</i> (N, M, D)
5	rs2764766	127213625	1.94×10^{-8}	<i>LINC01184</i> (N)
6	rs6903570	64866857	2.70×10^{-8}	<i>EYS</i> (N), <i>PHF3</i> (D), <i>PTP4A1</i> (D)
6	rs160593	105466332	7.69×10^{-9}	<i>HACE1</i> (Q), <i>LIN28B</i> (N, Q)
6	rs7757969	112132032	4.82×10^{-8}	<i>FYN</i> (N, Q)
6	rs4479915	165075601	4.82×10^{-9}	<i>C6ORF118</i> (N)
7	rs323167	78336677	4.47×10^{-8}	<i>MAGI2</i> (N, D)
7	rs11534004	113467444	1.71×10^{-8}	<i>PPP1R3A</i> (N, M)
8	rs17687067	17036201	3.39×10^{-12}	<i>MTMR7</i> (Q), <i>VPS37A</i> (Q), <i>ZDHHC2</i> (N, D, Q)
8	rs73219805	26272768	1.94×10^{-11}	<i>BNIP3L</i> (N, D), <i>PPP2R2A</i> (D), <i>SDAD1P1</i> (Q)
10	rs111364339	64857872	5.37×10^{-9}	<i>JMJD1C</i> (D), <i>NRBF2</i> (N)
12	rs28607014	117708611	1.75×10^{-8}	<i>NOS1</i> (N)
14	rs10148671	29469373	4.46×10^{-8}	<i>LINC01551</i> (N)
14	rs2383377	33257914	2.36×10^{-8}	<i>AKAP6</i> (N, D), <i>NPAS3</i> (D)
14	rs8012642	84669481	4.66×10^{-8}	<i>FLRT2</i> (N)
15	rs783540	83254708	3.05×10^{-8}	<i>AP3B2</i> (D, Q), <i>CPEB1</i> (N, Q)
15	rs758129	89900887	2.87×10^{-8}	<i>MIR9-3</i> (N), <i>POLG</i> (D), <i>RLBP1</i> (Q)
16	rs6500596	4470027	5.24×10^{-9}	<i>CDIP1</i> (Q), <i>CORO7</i> (N, D, Q), <i>DNAJA3</i> (M, Q), <i>NMRAL1</i> (Q, S)
16	rs8058130	64371163	4.77×10^{-8}	<i>CDH11</i> (N)
17	rs56007784	1290950	1.16×10^{-9}	<i>YWHAE</i> (N)
17	rs72843506	19946287	3.73×10^{-8}	<i>AKAP10</i> (D), <i>CCDC144CP</i> (Q), <i>SPECC1</i> (N, D, Q), <i>USP32P3</i> (Q)
17	rs35065479	55736735	2.31×10^{-8}	<i>TSPOAP1-AS1</i> (Q), <i>MSI2</i> (N)
18	rs56775891	77575613	1.85×10^{-8}	<i>KCNQ2</i> (N, Q, S)
18	rs28735056	77622879	4.60×10^{-10}	<i>KCNQ2</i> (N)

Genomic position is based on the UCSC hg19/NCBI build 37. ^aNotable genes are indicated as follows: gene nearest to the index SNP (N); schizophrenia-associated variant in strong LD ($r^2 \geq 0.8$) with a missense variant in the indicated gene (M); gene prioritized by DEPICT (D); gene with mRNA levels in *cis* genetic linkage with the index SNPs (Q); and gene prioritized by SMR analysis (S).

SNPs with no evidence of association ($P > 0.5$), there was no enrichment in coincident risk alleles across ancestry groups (Supplementary Table 7). We repeated this analysis by identifying the schizophrenia risk alleles at SNPs in the PGC2 data set and assessing concordance in the direction of the effect in the Chinese data set, and we found a very similar pattern (Supplementary Table 7). We concluded that there was a significant excess in directional concordance across ancestry groups for the SNPs with evidence of a schizophrenia association.

Potential biological mechanisms of the associated loci

To determine the likely causal genes of the schizophrenia-associated genetic loci, we considered each of the following to represent evidence supporting a gene's causality within a locus (Online Methods): (i) being the gene nearest the index SNP²³; (ii) containing a missense mutation and being in high LD ($r^2 > 0.8$) with the GWS SNPs²³; (iii) showing prioritization with DEPICT²⁴; (iv) being *cis*-acting expression quantitative trait loci (*cis*-eQTL) genes for the index SNPs^{23,25–28}; or (v) showing prioritization in summary-data-based Mendelian randomization (SMR) analysis²⁹. Using these criteria, we prioritized 247 genes from the schizophrenia risk loci and found that 85 had more than one line of supporting evidence (defined as 'prioritized candidate genes') (Supplementary Table 8). We first focused on those genes in the newly identified loci (Table 1). As expected, some of those genes were plausibly biologically relevant. The index SNP rs2247870 (NP_115495.3, p.Val587Ile) at 5q14.3 (GWS locus no. 37) is a missense variant in *ADGRV1* (also known as *GPR98*), which encodes

a member of the G-protein-coupled-receptor superfamily and is expressed in the central nervous system. Multiple lines of evidence suggest that G-protein-coupled receptors play critical roles in major psychiatric disorders (including schizophrenia) and their treatment³⁰. A variant in *GPR98* has been found to be associated with the response to antipsychotic treatment³¹. *FYN* (GWS locus no. 49) encodes a membrane-associated tyrosine kinase. *FYN* plays a critical role in neuronal apoptosis and is involved in brain development and synaptic transmission^{32,33}. Lower expression of *FYN* protein has been observed in the platelets of schizophrenic patients compared with controls³⁴. The results from whole-blood eQTL analysis²⁷ indicated that the schizophrenia risk allele identified in this study (rs7757969[C]) was correlated with a lower expression of *FYN* ($P = 1.71 \times 10^{-7}$, with a false discovery rate < 0.05 and in the credible interval covered by the 99% credible set), in agreement with previous findings. The estimate (b_{XY}) for the effect of gene expression on schizophrenia risk under the SMR analysis was -0.70 ($P_{SMR} = 7.55 \times 10^{-4}$). *MAGI2* (GWS locus no. 54) encodes a synaptic scaffolding molecule that is essential for the development and maintenance of synapses³⁵. Synaptic dysfunction has been suggested to play an important role in schizophrenia³⁶. Common variants in *MAGI2* have been found to be associated with cognitive impairment in people with schizophrenia³⁷. Although it is currently difficult to pinpoint a causal gene that is responsible for a given locus, the prioritized genes may be considered as favorable candidates for further research to unravel the plausible biological mechanisms underlying the associations.

Improved fine-mapping resolution at the associated loci

We sought to refine the localization of likely functional variants in the schizophrenia-associated loci by using a previously published approach^{38,39}. We derived Bayesian credibility sets in different data sets and evaluated the evidence for improved fine-mapping resolution through transancestry meta-analysis. For the 99% credible SNP sets, the transancestry data set produced the smallest spanned regions for ~80% ($n = 88$) of the tested loci (**Supplementary Table 9**), including 11 loci with a spanned region less than 30 kb. At 53 of the 88 loci, the number of genes that overlapped with the transancestry interval defined by a credible set was two or fewer. Of those overlapped genes mapping to the credible intervals in the 53 loci, 49.1% was in the list of prioritized candidate genes, whereas the proportion was 12.4% and 5.1% for the analysis of PGC2 and Chinese data, respectively.

We also conducted fine-mapping analysis with PAINTOR by leveraging the functional annotation data and LD information in multi-ancestry cohorts^{40–42}. We integrated the primary functional categories (coding, UTR, promoter, DNase-hypersensitivity site, intronic and intergenic) proposed by Gusev *et al.*⁴¹. A total of 62 variants achieved a posterior probability of >0.80 in at least one of the single-population and transancestry analyses (**Supplementary Table 10**). Of them, 38 variants had a higher posterior probability in the transethnic analysis than in the single-population analyses, including an additional 16 variants that achieved a transancestry posterior probability of >0.80 but had a posterior probability <0.80 in the single-population analyses. Eleven (68.8%) of these 16 variants had at least one hit in the selected eQTL studies in HaploReg v4.1 (ref. 23 and **Supplementary Table 11**). For example, at GWS locus no. 80, rs12541 with a posterior probability of 0.926 (**Supplementary Fig. 7a**) is in the UTR region of *ESAM* and correlated with its expression in whole blood ($P = 3.62 \times 10^{-8}$ and in the 99% credible-set interval)²⁷. A further example is GWS locus no. 103, rs3814883, which had a posterior probability of 0.911 (**Supplementary Fig. 7b**) and is a synonymous variant of *TAOK2* and also an eQTL SNP for several genes in different tissues²³ (**Supplementary Table 12**). It might also be correlated with the expression of *SEZ6L2* in the brain cerebellum and frontal cortex ($P = 2.37 \times 10^{-8}$ and 5.03×10^{-8} , respectively)²⁶. *TAOK2* has been found to affect basal-dendrite development in cortical neurons⁴³. *SEZ6L2* has been found to be a Cathepsin D transport receptor involved in neurite outgrowth⁴⁴. To further explore the regulatory nature in the context of the cell-type-specific epigenome, we also integrated the reference epigenomes of seven highlighted marks for 127 human tissues and cell types produced by the Roadmap Epigenomics Project⁴⁵ (Online Methods). Of the top 100 enriched cell-type-specific epigenomic annotations for schizophrenia associations in the current and PGC2 analyses⁷, over 40 were related to the brain and nervous system (**Supplementary Table 13**). In the further PAINTOR fine-mapping analyses with the top 100 epigenomic annotations, many SNPs had higher posterior probabilities, some of which increased to a value >0.80 , thus indicating potential biologically relevant cell types for these associations (**Supplementary Table 14**). For example, rs6670165, a candidate causal SNP at GWS locus no. 7, mapped to enhancers and promoters active in several brain regions. The identification of these SNPs suggested an important benefit of the transancestry fine-mapping signal in functional annotation data. However, 14 variants had a posterior probability >0.80 in the single-population analyses, which decreased to <0.8 in the transancestry analysis (**Supplementary Table 10**).

Biological pathways and gene sets

To identify pathways and gene sets in the transancestry meta-analysis, we performed an enrichment analysis with MAGMA⁴⁶. We identified

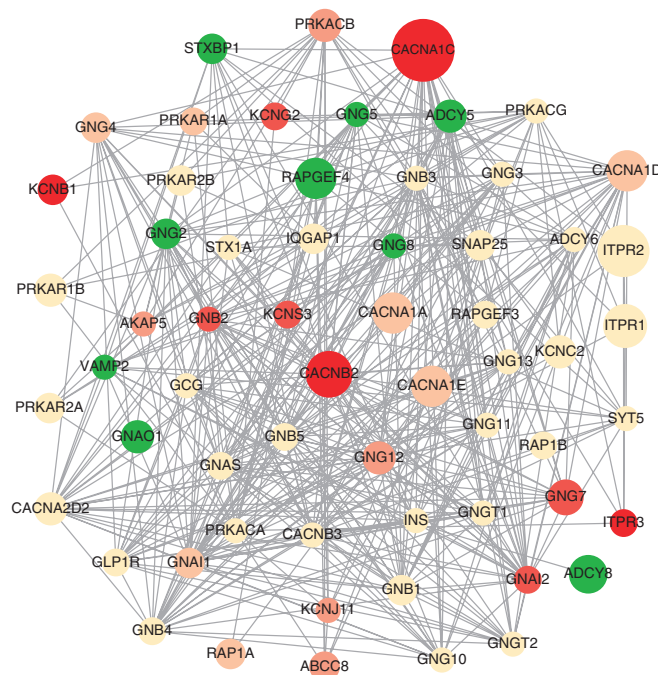


Figure 3 Interaction network of the schizophrenia-associated pathway 'glucagon-like peptide-1 regulates insulin secretion'. The network shows functional interactions for the genes in the pathway 'glucagon-like peptide-1 regulates insulin secretion' from the Reactome database. Each node represents a gene, and each edge represents a functional interaction. The node size corresponds to the gene size. The node color corresponds to the significance of the gene on the basis of the MAGMA analysis, and the green-to-red gradient corresponds to nonsignificance to high significance.

one gene set, 'regulation of insulin secretion by glucagon-like peptide 1' (from the Reactome database) that was significantly enriched (MAGMA competitive $P = 5.14 \times 10^{-7}$; **Fig. 3**). The MAGMA pathway analysis also highlighted several other pathways. Two of the previously highlighted schizophrenia-associated pathways, 'postsynaptic density'⁴⁷ and 'voltage-gated calcium channel complex'⁷, also ranked highly in our analysis, with P values of 9.01×10^{-4} and 1.32×10^{-3} , respectively (**Supplementary Table 15**).

Polygenic risk-score profiling

Polygenic scoring analyses have been proposed to predict the case-control status in a target data set, on the basis of the results from a training GWAS⁴. To assess the overlap between the common-variant signal in the European and Chinese populations and to provide estimates of the proportions of variance additionally explained by the Chinese sample, we conducted a polygenic scoring analysis. We randomly selected approximately 1,000 schizophrenia cases and 1,000 controls from the Chinese sample as the target sample and used four training data sets: (i) the PGC2 European-only data set (EUR49); (ii) the full PGC2 data set; (iii) the Chinese sample, excluding the target sample; and (iv) the Chinese plus PGC2 combined data set (**Fig. 4**). The risk-profile SNPs (P thresholds (P_T) = 5×10^{-8} , 1×10^{-6} , 0.0001, 0.001, 0.01, 0.05, 0.1, 0.2, 0.5 and 1) from the European-only data set alone explained approximately 1.11% to 2.34% of the variance in the case-control status of the Chinese sample on the liability scale⁴⁸ (assuming a population risk of 0.01). When the Asian samples were included, the PGC2 data set explained approximately 1.52% to 3.51% of the variance. The Chinese data set alone explained approxi-

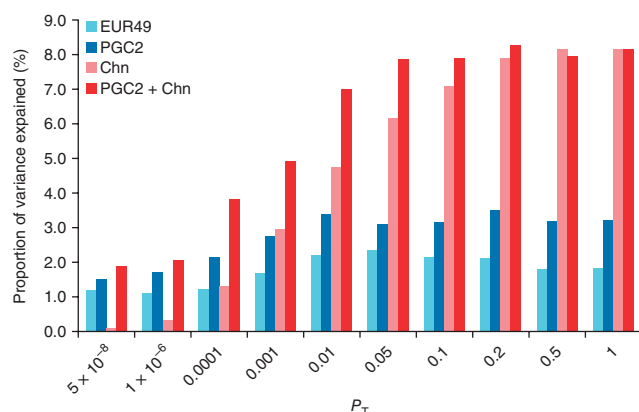


Figure 4 Polygenic risk-score profiling analysis. Polygenic risk-score profiling analysis using approximately 1,000 randomly selected schizophrenia cases and 1,000 controls from the Chinese sample as a target and deriving risk alleles from three training data sets: the PGC2 European-only (EUR49) data set (light blue); the full PGC2 data set (blue); the Chinese (Chn) sample excluding the target sample (light red); and the Chinese and PGC2 data sets combined (red). The x axis shows ten P_T -value thresholds ($P_T = 5 \times 10^{-8}$, 1×10^{-6} , 0.0001, 0.001, 0.01, 0.05, 0.1, 0.2, 0.5 and 1). The y axis is the estimate of the proportion of variance explained on the liability scale, which is converted from Nagelkerke's pseudo R^2 (computed by comparison of a full model including covariates and polygenic risk scores to a reduced model including covariates only).

mately 0.10% to 8.15% of the variance. In almost all situations, the combined data set (PGC2 plus Chinese) explained larger proportions of the variance (approximately 1.89% to 8.28%). For $P_T = 5 \times 10^{-8}$, the proportion of the explained variance increased from 1.20% (EUR49), 1.52% (PGC2) and 0.10% (Chinese) to 1.89% (PGC2 plus Chinese); for $P_T = 0.05$, it is increased from 2.34% (EUR49), 3.09% (PGC2) and 6.16% (Chinese) to 7.86% (PGC2 plus Chinese). To evaluate the increased variance explained by the newly identified GWS loci, we performed additional polygenic risk-score profiling trained on the full data set (GWAS excluding the target sample and with replication) and restricted to the newly identified loci with the Chinese sample included. These novel loci explained 1.34% of the variance, 30% of which was contributed by the loci from the Chinese-only analysis.

Correlations between two psychiatric disorders in the Chinese sample

Strong evidence of a shared genetic etiology between schizophrenia and other psychiatric disorders (such as bipolar disorder and major depressive disorder) has been observed in European samples^{49,50}. The degree of shared variation across psychiatric disorders in the Chinese population has been unclear. We estimated the genetic correlation between schizophrenia and major depressive disorder, two diseases for which Chinese GWAS data with large sample sizes are available, by using LD-score regression¹⁴. We observed a statistically significant genetic correlation between schizophrenia and major depressive disorder in the Chinese sample ($r_g = 0.43$, s.e. = 0.08, LD-score regression $P = 5.87 \times 10^{-8}$), in agreement with findings ($r_g = \sim 0.40$) in the European samples⁴⁹.

DISCUSSION

In the large GWAS analysis of schizophrenia in subjects of Chinese ancestry, we identified seven GWS loci, four of which were novel. In general, alleles identified as being associated at subthreshold levels of significance in the Chinese data set were also enriched in schizophrenia cases in the GWAS from PGC2, thus supporting the validity of combin-

ing the two data sets. The transancestry meta-analyses of the Chinese and PGC2 data identified 109 GWS risk loci, three of which were GWS in the Chinese-only analysis. Our analyses confirmed most of the previously identified schizophrenia loci and identified 30 novel loci.

We observed a significant excess in the directional consistency of schizophrenia risk alleles across ancestry groups, even at SNPs demonstrating only weak evidence of an association. These findings indicated that most schizophrenia risk loci were shared across these two ancestral populations, and transancestry meta-analysis provided a powerful means for identifying new loci and narrowing the association intervals. Polygenic scoring analysis also demonstrated notable increases in the explained variance in case-control status (PGC2-plus-Chinese training to Chinese target compared with PGC2 to Chinese target or Chinese training to Chinese target). However, this analysis also suggested that variants identified in European samples partially explained the genetic variance of schizophrenia in Chinese populations. Notably, estimates of the proportion of explained variance in liability were lower than those in European populations⁷, similarly to previous reports on transethnic analyses^{4,51}. Such lower estimates might be a result of differences in the allele frequencies and LD patterns between different populations⁴.

It has been suggested that there are also population-specific risk alleles for schizophrenia⁶ and that, if so, cross-ancestry analyses might have less power than that of studies of individuals with a recent shared ancestry. We found that some GWS loci in the PGC2 report were not GWS in the PGC2-plus-Chinese combined analysis. Moreover, most of the GWS SNPs identified in the analysis of Chinese subjects showed strong heterogeneity only across ancestries, though three of them achieved nominal significance with the same sign in the PGC2 data set. Another SNP fell within the previous PGC2-identified locus, but it was rare (MAF < 1%) in European populations. Thus, further transancestry fine-mapping, by leveraging the differences in the LD structure among diverse populations, may be an efficient approach to identify the causal variants underlying such associations and may also distinguish population-specific loci. Indeed, we also observed considerable improvements in the fine-mapping resolution at several susceptibility loci.

Our use of fine-mapping tools and functional annotations to analyze schizophrenia-associated loci identified numerous candidate genes with several lines of supporting evidence, including genes that have previously been implicated in schizophrenia (for example, *FYN* and *MAGI2*) and novel genes (for example, *EMX1* and *BNIP3L*) within the novel loci. Moreover, pathway analyses highlighted several pathways that contribute to schizophrenia pathogenesis, including previously described pathways (the voltage-gated calcium-channel pathway and postsynaptic density) and a new pathway (regulation of insulin secretion by glucagon-like peptide 1). The latter has not been highlighted in previous genetic studies of schizophrenia, but evidence from other investigation types has linked insulin signaling to the pathophysiology of schizophrenia. Previous epidemiological data have suggested that individuals with schizophrenia, compared with the general population or healthy controls, have a higher prevalence of metabolic syndrome^{52,53}. Moreover, high prevalence rates of impaired glucose metabolism have been observed in drug-naïve patients with schizophrenia⁵⁴. A proteomic analysis has shown that levels of several proteins involved in energy metabolism are altered in the brains of schizophrenic people⁵⁵. Our results provided further support for a role for insulin-related energy metabolism in the etiology of schizophrenia.

In summary, the Chinese ($n = 36,180$) and multi-ancestry ($n = 118,495$) GWAS meta-analysis and follow-up replication studies identified

113 GWS risk loci for schizophrenia, 30 of which are novel. Our results demonstrated added value from transancestry meta-analysis for fine-mapping of loci associated with schizophrenia and highlighted the existence of shared genetic risk across populations. In addition to confirming known genetic architectures, our comprehensive analyses provide further biological insights into the etiology of schizophrenia, thus potentially facilitating further mechanistic studies to assess the pathogenesis of this complex disorder.

URLs. PGC, <http://pgc.unc.edu/>; EIGENSTRAT, <https://github.com/DReichLab/EIG/tree/master/EIGENSTRAT/>; SHAPEIT, https://mathgen.stats.ox.ac.uk/genetics_software/shapeit/shapeit.html; IMPUTE2, http://mathgen.stats.ox.ac.uk/impute/impute_v2.html; 1000 Genomes Project, <http://www.1000genomes.org/>; The NIH Roadmap Epigenomics Mapping Consortium, <http://www.roadmapepigenomics.org/>; HaploReg v4.1, http://archive.broadinstitute.org/mammals/haploreg/haploreg_v4.1.php/; PLINK, <https://www.cog-genomics.org/plink2/>; PubMed, <http://www.ncbi.nlm.nih.gov/pubmed/>; NHGRI-EBI GWAS Catalog, <https://www.ebi.ac.uk/gwas/>; UCSC, <http://genome.ucsc.edu/>; GeneCards, <http://www.genecards.org/>; LDSC, <https://github.com/bulik/ldsc/>; A. Price laboratory, <https://www.hsph.harvard.edu/alkes-price/software>.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the [online version of the paper](#).

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank all of the participants in the study and the international Psychiatric GWAS Consortium (PGC) for the large-scale data resources that made this research possible. We also appreciate H. Huang, B. Neale and M. Daly for their valuable suggestions for data analysis and manuscript organization. This work was supported by the 973 Program (2015CB559100 to Y.S.), the National Key R&D Program of China (2016YFC0903402 to Y.S. and Z.L., and 2016YFC1201701 to X.L.), the Natural Science Foundation of China (31325014 to Y.S., 81130022 to Y.S., 81421061 to L.H. and 81701321 to Z.L.), the Program of Shanghai Subject Chief Scientist (15XD1502200 to Y.S.), the National Program for Support of Top-Notch Young Professionals to Y.S., the Shanghai Key Laboratory of Psychotic Disorders (13dz2260500 to Y.X.), the 'Shu Guang' project supported by the Shanghai Municipal Education Commission and Shanghai Education Development Foundation (12SG17 to Y.S.), the China Postdoctoral Science Foundation (2016M590615 to Z.L.), the Shandong Postdoctoral Innovation Foundation (201601015 to Z.L.), the Qingdao Postdoctoral Application Research Project (2016048 to Z.L.), the Shanghai Hospital Development Center (SHDC12016115 to Y.X.), the US NIMH and NIDA (U01 MH109528 to P.F.S. and U01 MH1095320 to P.F.S.), and the Swedish Research Council (Vetenskapsrådet, award D0886501 to P.F.S.).

AUTHOR CONTRIBUTIONS

Y.S. conceived and designed the experiments, and supervised all aspects of the work; J.C., Y.X., L.H., D.Z., W.Y., P.W., P.Y., B. Liu, W.S., Q.X., W.J., G.F., Q.Y., C.L. and X.L. performed sample collection and phenotyping; J.C., H.Y., J.Z., B.C., Y.L., J.W., J.J., M.W., Q.W., Z.W., Wenjin Li, K.L., F.H., J.Z., G.H., Weidong Li, C.W. and B. Li performed the experiments and data management; Z.L., H.Y., Z.S., J.S., S.R., P.F.S. and M.C.O'D. performed bioinformatics and statistical analyses; Y.S. and Z.L. interpreted the main findings; Y.S. and Z.L. drafted the manuscript; Y.S., L.H., Z.L., Y.X., X.L. and P.F.S. obtained the funding support; all authors revised and approved the final manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

- Knapp, M., Mangalore, R. & Simon, J. The global costs of schizophrenia. *Schizophr. Bull.* **30**, 279–293 (2004).
- Montgomery, W. *et al.* The personal, societal, and economic burden of schizophrenia in the People's Republic of China: implications for antipsychotic therapy. *Clinicoecon. Outcomes Res.* **5**, 407–418 (2013).
- Burmeister, M., McInnis, M.G. & Zöllner, S. Psychiatric genetics: progress amid controversy. *Nat. Rev. Genet.* **9**, 527–540 (2008).
- Purcell, S.M. *et al.* Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748–752 (2009).
- Ripke, S. *et al.* Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nat. Genet.* **45**, 1150–1159 (2013).
- de Candia, T.R. *et al.* Additive genetic variation in schizophrenia risk is shared by populations of African and European descent. *Am. J. Hum. Genet.* **93**, 463–470 (2013).
- Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).
- Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium. Genome-wide association study identifies five new schizophrenia loci. *Nat. Genet.* **43**, 969–976 (2011).
- Shi, J. *et al.* Common variants on chromosome 6p22.1 are associated with schizophrenia. *Nature* **460**, 753–757 (2009).
- Shi, Y. *et al.* Common variants on 8p12 and 1q24.2 confer risk of schizophrenia. *Nat. Genet.* **43**, 1224–1227 (2011).
- Stefansson, H. *et al.* Common variants conferring risk of schizophrenia. *Nature* **460**, 744–747 (2009).
- Steinberg, S. *et al.* Common variants at VRK2 and TCF4 conferring risk of schizophrenia. *Hum. Mol. Genet.* **20**, 4076–4081 (2011).
- Yue, W.-H. *et al.* Genome-wide association study identifies a susceptibility locus for schizophrenia in Han Chinese at 11p11.2. *Nat. Genet.* **43**, 1228–1231 (2011).
- Bulik-Sullivan, B.K. *et al.* LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
- Pocklington, A.J., O'Donovan, M. & Owen, M.J. The synapse in schizophrenia. *Eur. J. Neurosci.* **39**, 1059–1067 (2014).
- Lango Allen, H. *et al.* Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467**, 832–838 (2010).
- Winkler, T.W. *et al.* Quality control and conduct of genome-wide association meta-analyses. *Nat. Protoc.* **9**, 1192–1212 (2014).
- Yang, J. *et al.* Genomic inflation factors under polygenic inheritance. *Eur. J. Hum. Genet.* **19**, 807–812 (2011).
- Yu, H. *et al.* Common variants on 2p16.1, 6p22.1 and 10q24.32 are associated with schizophrenia in Han Chinese population. *Mol. Psychiatry* **22**, 954–960 (2017).
- Han, B. & Eskin, E. Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. *Am. J. Hum. Genet.* **88**, 586–598 (2011).
- Li, Z. *et al.* Loci with genome-wide associations with schizophrenia in the Han Chinese population. *Br. J. Psychiatry* **207**, 490–494 (2015).
- Nyholt, D.R. SECA: SNP effect concordance analysis using genome-wide association summary results. *Bioinformatics* **30**, 2086–2088 (2014).
- Ward, L.D. & Kellis, M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* **40**, D930–D934 (2012).
- Pers, T.H. *et al.* Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Commun.* **6**, 5890 (2015).
- GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
- Gibbs, J.R. *et al.* Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. *PLoS Genet.* **6**, e1000952 (2010).
- Westra, H.-J. *et al.* Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat. Genet.* **45**, 1238–1243 (2013).
- Fromer, M. *et al.* Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nat. Neurosci.* **19**, 1442–1453 (2016).
- Zhu, Z. *et al.* Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **48**, 481–487 (2016).
- Catapano, L.A. & Manji, H.K. G protein-coupled receptors in major psychiatric disorders. *Biochim. Biophys. Acta* **1768**, 976–993 (2007).
- Adkins, D.E. *et al.* Genomewide pharmacogenomic study of metabolic side effects to antipsychotic drugs. *Mol. Psychiatry* **16**, 321–332 (2011).
- Ali, D.W. & Salter, M.W. NMDA receptor regulation by Src kinase signalling in excitatory synaptic transmission and plasticity. *Curr. Opin. Neurobiol.* **11**, 336–342 (2001).
- Du, C.P., Tan, R. & Hou, X.Y. Fyn kinases play a critical role in neuronal apoptosis induced by oxygen and glucose deprivation or amyloid- β peptide treatment. *CNS Neurosci. Ther.* **18**, 754–761 (2012).
- Hattori, K. *et al.* Decreased expression of Fyn protein and disbalanced alternative splicing patterns in platelets from patients with schizophrenia. *Psychiatry Res.* **168**, 119–128 (2009).
- Bauß, K. *et al.* Phosphorylation of the Usher syndrome 1G protein SANS controls Magi2-mediated endocytosis. *Hum. Mol. Genet.* **23**, 3923–3942 (2014).

36. Calabrese, F., Riva, M.A. & Molteni, R. Synaptic alterations associated with depression and schizophrenia: potential as a therapeutic target. *Expert Opin. Ther. Targets* **20**, 1195–1207 (2016).
37. Koide, T. *et al.* Common variants in MAGI2 gene are associated with increased risk for cognitive impairment in schizophrenic patients. *PLoS One* **7**, e36836 (2012).
38. Gaulton, K.J. *et al.* Genetic fine mapping and genomic annotation defines causal mechanisms at type 2 diabetes susceptibility loci. *Nat. Genet.* **47**, 1415–1425 (2015).
39. Morris, D.L. *et al.* Genome-wide association meta-analysis in Chinese and European individuals identifies ten new loci associated with systemic lupus erythematosus. *Nat. Genet.* **48**, 940–946 (2016).
40. Farh, K.K.-H. *et al.* Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **518**, 337–343 (2015).
41. Gusev, A. *et al.* Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. *Am. J. Hum. Genet.* **95**, 535–552 (2014).
42. Kichaev, G. & Pasaniuc, B. Leveraging functional-annotation data in trans-ethnic fine-mapping studies. *Am. J. Hum. Genet.* **97**, 260–271 (2015).
43. de Anda, F.C. *et al.* Autism spectrum disorder susceptibility gene TAOX2 affects basal dendrite formation in the neocortex. *Nat. Neurosci.* **15**, 1022–1031 (2012).
44. Boonen, M. *et al.* Cathepsin D and its newly identified transport receptor SEZ6L2 can modulate neurite outgrowth. *J. Cell Sci.* **129**, 557–568 (2016).
45. Kundaje, A. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
46. de Leeuw, C.A., Mooij, J.M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLOS Comput. Biol.* **11**, e1004219 (2015).
47. Network and Pathway Analysis Subgroup of Psychiatric Genomics Consortium. Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. *Nat. Neurosci.* **18**, 199–209 (2015).
48. Lee, S.H., Goddard, M.E., Wray, N.R. & Visscher, P.M. A better coefficient of determination for genetic profile analysis. *Genet. Epidemiol.* **36**, 214–224 (2012).
49. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
50. Lee, S.H. *et al.* Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat. Genet.* **45**, 984–994 (2013).
51. Ikeda, M. *et al.* Genome-wide association study of schizophrenia in a Japanese population. *Biol. Psychiatry* **69**, 472–478 (2011).
52. Citrome, L. & Volavka, J. Consensus development conference on antipsychotic drugs and obesity and diabetes: response to consensus statement. *Diabetes Care* **27**, 2087–2088, author reply 2089–2090 (2004).
53. Mitchell, A.J. *et al.* Prevalence of metabolic syndrome and metabolic abnormalities in schizophrenia and related disorders: a systematic review and meta-analysis. *Schizophr. Bull.* **39**, 306–318 (2013).
54. Steiner, J. *et al.* Immune system and glucose metabolism interaction in schizophrenia: a chicken-egg dilemma. *Prog. Neuropsychopharmacol. Biol. Psychiatry* **48**, 287–294 (2014).
55. Gottschalk, M.G., Wesseling, H., Guest, P.C. & Bahn, S. Proteomic enrichment analysis of psychotic and affective disorders reveals common signatures in presynaptic glutamatergic signaling and energy metabolism. *Int. J. Neuropsychopharmacol.* **18**, pyu019 (2014).

ONLINE METHODS

Recruitment of research subjects. As in our previous study¹⁰, all cases of Chinese ancestry were inpatients or outpatients with a history of more than 2 years of schizophrenia, who were recruited from mental-health centers in China, interviewed by two independent psychiatrists and diagnosed according to Diagnostic and Statistical Manual of Mental Disorders (DSM)-IV criteria. All cases met the following two 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 effects. The controls were randomly selected from Chinese volunteers (from hospitals and a community survey) who were asked to reply to a written invitation to evaluate their medical histories. Lists of potential control subjects were screened for suitability as volunteers by excluding subjects with major mental illnesses. All participants provided written informed consent. The study was approved by the Ethics Committee of Human Genetic Resources at the Bio-X Institutes of Shanghai Jiao Tong University, in accordance with the tenets of the Declaration of Helsinki. We confirm that our study is compliant with the Guidance of the Ministry of Science and Technology (MOST) for the Review and Approval of Human Genetic Resources.

Genotyping, quality control and genotype imputation of the Chinese GWAS data. Several different genome-wide genotyping platforms were used in this study: Affymetrix Genome-Wide Human SNP Array 6.0 (SNP6.0), Affymetrix Axiom Genome-Wide CHB1 Array Plate and Illumina 1M Array.

For the SNP6.0 chips, the genotype calls were generated together by using Affymetrix Axiom Analysis according to the Best Practices Workflow for SNP6.0. Sample QC filtering of the GWAS data was first performed by excluding arrays with Contrast QC measurements (a metric developed by Affymetrix for SNP6.0 QC, $n = 197$) that were <0.4 . Step 1 genotyping was run on all CEL files passing QC over a subset of 20,000 SNPs, and samples with a call rate $\leq 97\%$ were excluded ($n = 285$). The remaining samples were used for step 2 genotyping analysis. SNP polisher was then used for SNP QC, and the SNPs in the recommended categories (PolyHighRes, MonoHighRes, NoMinorHom and Hemizygous) were retained. Sex was established via genotyping and evaluated for each of the subjects, and samples with inconsistent sex (compared with the sample record) were removed ($n = 79$). Heterozygosity rates were calculated with the intent of removing deviations that exceeded 6 s.d. from the mean ($n = 0$). PLINK's identity-by-descent analysis was used to detect cryptic relatedness⁵⁶ (URLs). When a pair of individuals had $PI_{\text{HAT}} > 0.2$, the member of the pair with the lower call rate was excluded from the analysis ($n = 259$). SNPs with call rates $< 97\%$ ($n = 28,040$), $MAF < 1\%$ ($n = 185,439$) or significant deviation from Hardy-Weinberg equilibrium (HWE) in controls ($HWE P \leq 1 \times 10^{-6}$, $n = 20,344$) were excluded. We also excluded population outliers on the basis of PCA. After application of quality-control criteria, a set of 590,413 SNPs for 14,645 individuals was generated for genotype imputation.

For the CHB1 chips, the genotype calls were generated together according to the Axiom Genotyping Solution Data Analysis Guide. Briefly, arrays with dish QC (DQC), a single-sample metric developed by Affymetrix for Axiom QC) values < 0.82 were first excluded ($n = 181$). Samples that surpassed the DQC values were used for genotype calling with a subset of probe sets. Samples with a call rate $< 97\%$ or in a nonpassing plate (an average call rate of passing samples $< 98.5\%$) were also excluded ($n = 276$). The post-QC samples were then clustered, and genotype calls were produced with the Axiom Genotyping Algorithm v1 (Axiom GT1). SNP QC was also executed with the SNP polisher procedure, and the SNPs in the recommended categories were retained. Verification procedures for sex, relatedness and PCA outliers were also conducted in sample QC as described above ($n = 289$). SNPs with call rates $< 97\%$ ($n = 56,735$), $MAF < 1\%$ ($n = 206$) or significant deviations from HWE in controls ($HWE P \leq 1 \times 10^{-6}$, $n = 18,849$) were excluded. After application of QC criteria, a set of 555,058 SNPs for 9,580 individuals was generated for genotype imputation.

For Illumina 1M chips, SNP genotypes were generated from normalized bead intensity data with Genome Studio. Samples with a call rate $< 97\%$ were excluded ($n = 35$). Regular sample QC procedures for parameters including sex, relatedness, heterozygosity rate and PCA outlier checking, were performed

as described above ($n = 231$). SNPs with call rates $< 97\%$ ($n = 35,743$), $MAF < 1\%$ ($n = 89,032$) or $HWE P \leq 1 \times 10^{-6}$ ($n = 954$) were excluded. After application of QC criteria, a set of 716,466 SNPs for 1,823 individuals was generated for genotype imputation.

For each GWAS data set, the entire set was imputed together as follows: the genotypes were phased with SHAPEIT (URLs)^{57,58} for each chromosome, and imputation was performed for each 5-Mb chromosome interval with IMPUTE2 (URLs)⁵⁹. The haplotypes derived from the 1000 Genomes Project Phase 1 (release v3, URLs) were used as reference data⁶⁰. Because two genotyping platforms were used for GWAS set 3, we used two phased reference panels in this special case, as proposed by Howie *et al.*⁵⁹. For each platform, the prephased data from the other platform were used as the second reference panel. The variants with $INFO > 0.8$, $MAF > 0.01$, a call rate $\geq 97\%$ and $HWE P \geq 1 \times 10^{-6}$ in the controls were saved for further analysis. Those present in at least two data sets were saved for the meta-analysis. A set of 5,107,227 genetic variants for 7,699 cases and 18,327 controls remained in the final analysis.

PGC2 GWAS data set. The PGC2 GWAS data set⁷ comprised 49 case-control samples (34,241 cases and 45,604 controls) and three family-based samples (1,235 parent-affected offspring trios). All of the samples were from subjects of European ancestry, excluding three case-control samples from subjects of East Asian ancestry (1,836 cases and 3,383 controls). The summary results for the PGC2 data set and European only data set (EUR49) were downloaded from the PGC website (URLs).

Replication data set. The replication sample consisted of 4,384 cases and 5,770 controls of Han Chinese ancestry. More details of the general characteristics and genotyping have been presented in our previous research¹⁹. For the Chinese-only analyses, the independent SNPs with $P < 1 \times 10^{-5}$ in the Chinese GWAS analysis of pre- or postcorrection with the inflation factor were selected. For the transancestry analysis, the independent SNPs with $P < 5 \times 10^{-7}$ in the Chinese (pre- or postcorrection) and PGC2 GWAS meta-analyses were selected. The precorrection data sets were used only for including more candidate SNPs for replication. All the association results in this article were based on the postcorrection data sets, wherein the global inflations were controlled. A total of 295 SNPs were analyzed in the Chinese replication analysis.

Power calculations. Power calculations were performed with the GAS Power Calculator⁶¹ with a range of genotype relative risks and disease-allele frequencies, assuming a population prevalence of 0.01 and a significance level of 5×10^{-8} . For the Chinese-only ($n = 36,180$) and transancestry ($n = 118,495$) analyses, we had adequate power ($> 80\%$) to detect variants with low risk-allele frequencies (RAFs) of 0.03 with genotypic relative risks of 1.318 and 1.161, respectively. This sample size in Chinese-only analyses was large enough to achieve adequate power for risk variants with genotypic relative risks of 1.150 and RAFs of 0.14 to 0.85, and the transancestry analyses achieved adequate power for risk variants with 1.075 and RAFs of 0.15 to 0.84.

Statistical methods and bioinformatics analysis. Population substructure was evaluated through a PCA with EIGENSTRAT software (URLs), on the basis of LD-pruned autosomal SNP genotypes^{62,63}. Two rounds PCA were performed. One round with samples from the HapMap Project phase 3 (HapMap3) was performed to identify admixed samples, and the other round was performed for each subset of cases and controls, wherein individual outliers (> 6 s.d. from the mean on any one of the top ten PCs) were identified and removed for five iterations, and final PCs reflecting subtle ancestry information for each sample were generated for further correction. In the Chinese GWAS stage, the association was analyzed for subsets by using a logistic regression model involving covariates for PCs to adjust for possible population stratification. We evaluated the effects of the 20 PCs on genome-wide test statistics to determine the PC inclusion in the final association analysis for each data set. In the Chinese replication stage, the associations between SNPs and schizophrenia risk were evaluated on the basis of logistic regression with SNPTTEST⁶⁴. The Higgins and Thompson I^2 index was used for assessing heterogeneity across data sets⁶⁵. Both fixed-effects-model and random-effects-model meta-analyses were used in this study. The variants with pronounced heterogeneity ($I^2 > 75\%$) were combined in a random-effects model in the transancestry meta-analysis¹⁷.

We assessed the genome-wide congruence of risk alleles across samples by using binomial sign tests that compared the direction of the effect sizes of independent SNPs between PGC2 and Chinese GWAS results. *P* values were generated under the null hypothesis ($H_0: P = 0.50$). The proportion of variance in liability to schizophrenia explained by the common SNPs was estimated by using genome-wide complex-trait analysis⁶⁶, and the PCs were included in the analysis as covariates. For each of the associated loci (except the eMHC region, owing to the complexity of this region⁷), we calculated an approximate Bayes factor per Wakefield, as well as the posterior probability of driving the association for each SNP within a 2-Mb window, and then created 99% credibility sets^{38,39,67}. We created credibility sets by using the Chinese, PGC2 (European) and combined data sets separately. We conducted the transancestry fine-mapping in the presence of functional information by using PAINTOR according to the suggested pipeline, as well as PGC2-only and Chinese-only analyses for comparison. The primary functional annotations for SNPs proposed by Gusev *et al.*⁴¹ were obtained from the A. Price laboratory website (URLs). The reference epigenomes of 127 human tissues and cell types⁴⁵ were obtained from the NIH Roadmap Epigenomics Mapping Consortium (URLs). We included seven highlighted epigenomic marks (H3K4me3, H3K4me1, H3K36me3, H3K27me3, H3K9me3, H3K27ac and H3K9ac)⁴⁵ in our analyses. Enrichment analyses of the schizophrenia associations in the current and PGC2 analyses with the epigenomic features were performed with the genomic regulatory elements and GWAS overlap algorithm (GREGOR)⁶⁸, and the top 100 enriched annotations were selected for further PAINTOR analyses. The online tool HaploReg²³ (v4.1; URLs) was used to explore the genes nearest to the index SNPs, and genes containing a missense mutation in high LD ($r^2 > 0.8$, on the basis of the 1000 Genomes Phase 1 CEU or ASI population for the LD calculation) with the GWS SNPs. The effects of GWS SNPs on expression in eQTL studies of different tissues (including blood and brain tissues^{25–27}) were extracted from the query results of HaploReg²³ and the CommonMind Consortium Knowledge Portal²⁸. A significant eQTL was reported as having a false discovery rate of 0.05 in the original studies^{25–28} and being located in the credible interval covered by the 99% credible set for the regulated gene for the data sets in which detailed results were available for establishing the credible sets^{25,27}. We used DEPICT²⁴ to identify the most likely causal genes for the schizophrenia-associated loci, on the basis of the functional similarity among genes from associated regions. We carried out SMR analysis²⁹ for the blood and brain tissue eQTL data sets^{25,27}, using the 1000 Genomes Project data as reference files. For the gene prioritization analysis at the GWS loci (excluding the eMHC region, owing to the complexity of this region²⁹), only probes with at least one *cis*-eQTL at $P < 5.0 \times 10^{-8}$ were considered for SMR analysis, and a significance threshold was set as $P_{SMR} < 5.20 \times 10^{-5}$ corresponding to a Bonferroni correction for 960 tests (960 probes with *cis*-eQTL at $P < 5.0 \times 10^{-8}$ across the GWS loci)²⁹. The heterogeneity in dependent instruments (HEIDI) test was also performed, and $P < 0.05$ was considered to indicate significant heterogeneity. The genes prioritized by the GWS index SNP or its high LD ($r^2 > 0.8$) proxies were listed. In addition, the SMR analysis was also performed for some specific SNPs and genes. Here, the *P*-value threshold for selecting eQTL was not applicable, and the details are shown in the results. We searched the published literature for these genes with respect to schizophrenia in PubMed (URLs) and the NHGRI-EBI GWAS Catalog (URLs), and we obtained additional functional evidence for these SNPs and genes from the published literature, the UCSC genome database (URLs) and GeneCards (URLs).

LD-score regression for Chinese GWAS data. We estimated Chinese LD scores from the Chinese subjects in the 1000 Genomes Project Phase 3, using the LD Score (LDSC; URLs) software package¹⁴. We used a window size of 1 cM to estimate LD scores, excluded singletons and did not set an r^2 cutoff. The LD-score regression intercept from the Chinese GWAS data was estimated according to application notes for real data from the LDSC developers¹⁴. As Bulik-Sullivan *et al.* have proposed¹⁴, correcting test statistics with the LD-score regression intercept is a robust way for controlling the confounding bias from inflation.

Correction was applied to the Chinese GWAS meta-analysis results by multiplying the standard errors by the square root of the correction factor¹⁶.

Polygenic scoring analysis. Approximately 1,000 cases and 1,000 controls from the Chinese sample were randomly selected as the target sample. Risk-profile SNPs ($P_T = 5 \times 10^{-8}$, 1×10^{-6} , 0.0001, 0.001, 0.01, 0.05, 0.1, 0.2, 0.5 and 1) from the training GWAS data sets (the PGC2 European-only (EUR49) and full data sets, the Chinese GWAS data set excluding the target sample and the Chinese plus PGC2 combined data set) were selected with the PLINK ‘--clumped’ function, and SNPs within 500 kb or with $r^2 \geq 0.1$ were discarded. The risk-profile SNPs were then used to generate scores for the target samples by using the PLINK ‘--score’ function. The case-control status was then predicted by logistic regression analysis of polygenic scores plus PC covariates. Nagelkerke’s R^2 was used for the full model, using the polygenic score plus the covariates minus R^2 for the covariates alone, thus yielding an estimate of the explained variance. The R^2 was then transformed into a liability scale⁴⁸, assuming a population prevalence of 1% for schizophrenia⁷.

Pathway analysis. MAGMA⁴⁶ was used to explore pathway-based associations in the genome-wide meta-analysis data set. An *F* test was used to compute the gene *P* value, and the gene *P* values and gene correlation matrix were then used for the gene-set analysis with a regression model⁴⁶. We defined gene boundaries 35 kb upstream and 10 kb downstream for assigning SNPs to a gene, as adopted in a recent psychiatric-disorder pathway analysis⁴⁷. Each gene was then assigned pathways in the Gene Ontology (GO), PANTHER, Ingenuity, Kyoto Encyclopedia of Genes and Genomes (KEGG), Reactome and BioCarta gene set databases⁶⁹. A total of 2,981 pathways or gene sets were used in this analysis.

Data availability. Summary statistics for the meta-analyses will be made available at <http://gwas.bio-x.cn/>. A Life Sciences Reporting Summary is available.

56. 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).
57. Delaneau, O., Marchini, J. & Zagury, J.-F. A linear complexity phasing method for thousands of genomes. *Nat. Methods* **9**, 179–181 (2011).
58. Delaneau, O., Zagury, J.-F. & Marchini, J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat. Methods* **10**, 5–6 (2013).
59. Howie, B.N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* **5**, e1000529 (2009).
60. 1000 Genomes Project Consortium. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
61. Skol, A.D., Scott, L.J., Abecasis, G.R. & Boehnke, M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat. Genet.* **38**, 209–213 (2006).
62. Price, A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
63. Patterson, N., Price, A.L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2**, e190 (2006).
64. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat. Genet.* **39**, 906–913 (2007).
65. Higgins, J.P.T., Thompson, S.G., Deeks, J.J. & Altman, D.G. Measuring inconsistency in meta-analyses. *Br. Med. J.* **327**, 557–560 (2003).
66. Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
67. Wakefield, J. Bayes factors for genome-wide association studies: comparison with *P*-values. *Genet. Epidemiol.* **33**, 79–86 (2009).
68. Schmidt, E.M. *et al.* GREGOR: evaluating global enrichment of trait-associated variants in epigenomic features using a systematic, data-driven approach. *Bioinformatics* **31**, 2601–2606 (2015).
69. Segre, A.V. *et al.* Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. *PLoS Genet.* **6**, e1001058 (2010).

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see [Reporting Life Sciences Research](#). For further information on Nature Research policies, including our [data availability policy](#), see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

► Experimental design

1. Sample size

Describe how sample size was determined.

No initial power analysis was done to determine the sample size. Our post hoc power analysis indicated that our sample size was large enough to achieve adequate power for detecting variants of low risk allele frequencies of 0.03 with genotypic relative risks of 1.161.

2. Data exclusions

Describe any data exclusions.

Typical quality control was performed for our GWAS data sets. Arrays with low quality data were excluded. Samples failed in the sex, relatedness, heterozygosity rate and PCA outlier checking procedures were also excluded.

3. Replication

Describe whether the experimental findings were reliably reproduced.

We performed Chinese and multi-ancestry GWAS meta-analyses and follow-up replication analyses, and the identified loci were reliably reproduced with genome-wide significant evidence.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

The samples were grouped by disease status.

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

The researchers were not blinded to group allocation.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a Confirmed

- | | | |
|-------------------------------------|-------------------------------------|--|
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | The <u>exact sample size</u> (<i>n</i>) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.) |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | A statement indicating how many times each experiment was replicated |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section) |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | A description of any assumptions or corrections, such as an adjustment for multiple comparisons |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | The test results (e.g. <i>P</i> values) given as exact values whenever possible and with confidence intervals noted |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | A clear description of statistics including <u>central tendency</u> (e.g. median, mean) and <u>variation</u> (e.g. standard deviation, interquartile range) |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | Clearly defined error bars |

See the web collection on [statistics for biologists](#) for further resources and guidance.

► Software

Policy information about [availability of computer code](#)

7. Software

Describe the software used to analyze the data in this study.

The URLs for the software used were provided.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* [guidance for providing algorithms and software for publication](#) provides further information on this topic.

► Materials and reagents

Policy information about [availability of materials](#)

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

The materials were commercially available.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

Not applicable.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

Not applicable.

b. Describe the method of cell line authentication used.

Not applicable.

c. Report whether the cell lines were tested for mycoplasma contamination.

Not applicable.

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by [ICLAC](#), provide a scientific rationale for their use.

Not applicable.

► Animals and human research participants

Policy information about [studies involving animals](#); when reporting animal research, follow the [ARRIVE guidelines](#)

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.

Not applicable.

Policy information about [studies involving human research participants](#)

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

For privacy concerns, we can't provide detailed information for the participants. These information were not used as covariates in our analysis.