

Common Variants on Xq28 Conferring Risk of Schizophrenia in Han Chinese

Emily H. M. Wong^{1,11}, Hon-Cheong So^{1,11}, Miaoxin Li^{1,2}, Quang Wang³, Amy W. Butler^{1,4}, Basil Paul¹, Hei-Man Wu¹, Tomy C. K. Hui¹, Siu-Chung Choi¹, Man-Ting So⁵, Maria-Mercè Garcia-Barcelo^{2,5}, Grainne M. McAlonan⁶, Eric Y. H. Chen¹, Eric F. C. Cheung⁷, Raymond C. K. Chan⁸, Shaun M. Purcell⁹, Stacey S. Cherny^{1,2,10}, Ronald R. L. Chen¹, Tao Li³, and Pak-Chung Sham^{*,1,2,10}

¹Department of Psychiatry, The University of Hong Kong, Hong Kong, China; ²Centre for Genomic Sciences, The University of Hong Kong, Hong Kong, China; ³The Mental Health Centre and Psychiatric Laboratory, West China Hospital, Sichuan University, Chengdu, Sichuan, China; ⁴MRC Social Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, King's College London, London, UK; ⁵Department of Surgery, The University of Hong Kong, Hong Kong, China; ⁶Department of Forensic and Neurodevelopmental Sciences, Institute of Psychiatry, King's College London, UK; ⁷Castle Peak Hospital, Hong Kong, China; ⁸Institute of Psychology, Chinese Academy of Sciences, Beijing, China; ⁹Division of Psychiatric Genomics, Mount Sinai School of Medicine, New York; ¹⁰State Key Laboratory in Brain and Cognitive Sciences, The University of Hong Kong, Hong Kong, China; ¹¹Co-first authors

*To whom correspondence should be addressed; 6/F The Hong Kong Jockey Club Building for Interdisciplinary Research, 5 Sassoon Road, The University of Hong Kong, Pokfulam, Hong Kong Special Administrative Region, China; tel: 2831 5425, fax: 2818 5653, e-mail: psham@hku.hk

Schizophrenia is a highly heritable, severe psychiatric disorder affecting approximately 1% of the world population. A substantial portion of heritability is still unexplained and the pathophysiology of schizophrenia remains to be elucidated. To identify more schizophrenia susceptibility loci, we performed a genome-wide association study (GWAS) on 498 patients with schizophrenia and 2025 controls from the Han Chinese population, and a follow-up study on 1027 cases and 1005 controls. In the follow-up study, we included 384 single nucleotide polymorphisms (SNPs) which were selected from the top hits in our GWAS (130 SNPs) and from previously implicated loci for schizophrenia based on the SZGene database, NHGRI GWAS Catalog, copy number variation studies, GWAS meta-analysis results from the international Psychiatric Genomics Consortium (PGC) and candidate genes from plausible biological pathways (254 SNPs).

Within the chromosomal region Xq28, SNP rs2269372 in *RENBP* achieved genome-wide significance with a combined P value of 3.98×10^{-8} (OR of allele A = 1.31). SNPs with suggestive P values were identified within 2 genes that have been previously implicated in schizophrenia, *MECP2* (rs2734647, $P_{\text{combined}} = 8.78 \times 10^{-7}$, OR = 1.28; rs2239464, $P_{\text{combined}} = 6.71 \times 10^{-6}$, OR = 1.26) and *ARHGAP4* (rs2269368, $P_{\text{combined}} = 4.74 \times 10^{-7}$, OR = 1.25). In addition, the patient sample in our follow-up study showed a significantly greater burden for pre-defined risk alleles based on the SNPs selected than the controls. This indicates the existence of schizophrenia susceptibility loci among the SNPs we selected. This also further supports multigenic

inheritance in schizophrenia. Our findings identified a new schizophrenia susceptibility locus on Xq28, which harbor the genes *RENBP*, *MECP2*, and *ARHGAP4*.

Key words: schizophrenia/genome-wide association study/Han Chinese/*MECP2*/*ARHGAP4*/*RENBP*

Introduction

Schizophrenia is a severe, highly heritable psychiatric disorder, affecting approximately 1% of the world population. It is characterized by delusions, hallucinations, disorganized thinking, and often social dysfunction. Considerable research efforts have been made to identify schizophrenia susceptibility genes. According to the Schizophrenia Research Forum database¹ (SZGene, as accessed on November 20, 2011), at least 1727 association studies covering 1008 candidate genes and 8788 polymorphisms have been performed on schizophrenia and related spectrum disorders. These studies range from candidate loci or pathway testing to genome-wide association studies (GWAS), with setting based on either single-nucleotide polymorphism (SNP) or copy number variation (CNV).

With the advancement in genotyping technology, interrogation of the entire genome for disease susceptibility variants has become feasible. Genome-wide association has proved to be a useful tool in the discovery of common variants for complex disorders. A number of GWAS on schizophrenia have been

published²⁻⁶ and included in public databases such as the NHGRI GWAS Catalog (<http://www.genome.gov/gwastudies/>) and the SZGene Database. However, the majority of these GWAS focus on populations of European ancestry, with the exception of 2 on the Han Chinese population.^{4,6} Although many common causal variants are likely to confer risk irrespective of ancestry, the frequency and linkage disequilibrium (LD) pattern of variants may differ across populations. Due to genetic heterogeneity, some susceptibility variants may be more readily discovered in one population than the others. The effect of a variant in different populations may also vary because of interaction with different genetic and environmental backgrounds.

To understand more about the genetic component of schizophrenia in the Han Chinese population, we have conducted a 2-stage schizophrenia study aiming to elucidate genetic variants that can confer risk of schizophrenia in Han Chinese. Our study consisted of a discovery phase GWAS and a follow-up study using an independent sample. The follow-up study was designed for genotyping SNPs selected based on the gene-based tests performed in our GWAS, and also genotyping loci previously implicated by other published studies. These loci included those reported in the SZGene Database, NHGRI GWAS Catalog (schizophrenia/bipolar disorder studies), CNV studies on schizophrenia, meta-analyses of GWAS results as provided by the International Psychiatric Genetics Consortium (PGC), and also genes in plausible biological pathways based on functional relevance. The inclusion of SNPs for the follow-up study based on our GWAS and external evidence provides a more comprehensive study of the genetics of schizophrenia. More information on SNP selection is provided in the supplementary materials and methods.

Materials and Methods

Subject Recruitment

The overall study was approved by the institutional review board (IRB) of The University of Hong Kong and West China Hospital of Sichuan University IRB. Subjects recruited were all of Han Chinese origin. Patients with schizophrenia were recruited from Castle Peak Hospital in Hong Kong (HK) and West China Hospital of Sichuan University in Sichuan, China and were assessed using the Structured Clinical Interview for DSM-IV (Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition) Patient Version (SCID-P). Included patients fulfilled the diagnostic criteria for schizophrenia as specified in DSM-IV.

Healthy controls were recruited from the local area and were screened for the lifetime absence of psychiatric illnesses by the SCID nonpatient version (SCID-NP). All the controls from HK were blood donors, while

controls from Sichuan were recruited from the community in Chengdu City, Sichuan Province, China by poster advertisement.

Discovery Phase. The case sample consisted of 400 patients with schizophrenia recruited from HK and 98 from Sichuan, a province in Southwest China. A total of 2071 control subjects were recruited from 5 different groups: (i) 800 subjects from a GWAS on bone mineral density in HK⁷; (ii) 1000 Taiwan Chinese control subjects provided by deCODE Genetics (Reykjavik, Iceland); (iii) 111 unrelated unaffected sibs from a family-based GWAS for hypertension in HK⁸; (iv) 100 control subjects from a GWAS for hepatocellular carcinoma in HK⁹; and (v) 60 healthy controls recruited from Sichuan.

Follow-up Study. The follow-up study used an independent sample of 1027 patients with schizophrenia (HK: $n = 69$; Sichuan: $n = 958$) and 1001 controls (HK: $n = 74$; Sichuan: $n = 931$). In addition, 71 affected offspring and 173 unaffected family members from 71 families in Sichuan were included. Each of these families had at least one affected offspring and 2 unaffected parents.

(1) Discovery Phase: GWAS

Genotyping and Quality Control

Genotyping for all HK and Sichuan samples was conducted by deCODE Genetics using the Illumina Human610-Quad BeadChip, while the Taiwan Chinese subjects were genotyped using the Illumina Human550-Quad Beadchip. Genotype data were subjected to standard quality control procedures (see supplementary information). Subjects were removed based on the following criteria: genotyping rate <95%, excessive level of heterozygosity, sample relatedness or duplication. SNPs were excluded if genotyping rate < 90%, minor allele frequency (MAF) < 0.01, or Hardy-Weinberg equilibrium (HWE, controls only) test P value < 1×10^{-6} .

Statistical Analysis

EIGENSOFT¹⁰ was used to detect population stratification in the GWAS dataset. The top 10 principal components (PCs) were extracted for subsequent statistical analyses. Genotype imputation was conducted using IMPUTE2¹¹, with the Chinese panels (CHB and CHD in phase 2 and phase 3) in the HapMap project as reference population. Imputed SNPs with an INFO score < 0.4 or MAF < 0.005 were removed. After imputation, SNPTEST¹¹ was used to perform statistical association tests for each individual SNP, under a logistic regression model with adjustment for sex and the top 10 PCs from EIGENSOFT.

Random inactivation of the X-chromosome occurs in females at early fetal development, such that only 1 allele

is expressed from each locus. In the absence of environmental factors or interaction with other loci, the hemizygous allelic effects of X-chromosome loci in males are expected to be similar to homozygous effects in females. Thus, for our analysis we double counted the allele on the X-chromosome in males. Genotypes of X-chromosome loci were coded as ["0" or "2"] for males and ["0," "1" or "2"] for females. This method of association analysis for chromosome X loci was previously suggested by Clayton.^{12,13} Analyses were also performed on male and female genotypes separately, and the overall association *P* values for X-chromosome SNPs were generated by combining the male and female association test results using the inverse weighted variance method.

After the SNP association tests, we proceeded to perform gene-based association tests using GATES (Gene-Based Association Test Using Extended Simes Procedure),¹⁴ a program implemented in KGG (Knowledge-based mining system for Genome-wide Genetic studies; <http://bioinfo.hku.hk/kggweb/>). GATES derives a gene-based *P* value by combining the association *P* values of SNPs within the same gene (coding sequences ± 5 kb flanking regions). The gene-based test accounts for gene size and LD pattern. Genes were categorized as candidate or noncandidate. For both candidate genes and noncandidate genes, significant association with schizophrenia was determined by FDR (false discovery rate) threshold of 0.55. The best SNPs from the significant candidate and noncandidate genes (ie, SNP with the best *P* value within the gene and having minor allele frequency > 0.01 in HapMap Chinese samples—CHB panel) were selected for genotyping in the follow-up study. SNPs that fall outside genic regions were also defined as significant according to the same FDR threshold and selected for genotyping in the subsequent stage. Using this approach, we prioritized 130 SNPs for the genotyping in the follow-up study.

Follow-up Study in Independent Sample

Selection of SNPs

In the second stage, follow-up study, we aimed to genotype SNPs prioritized based on our GWAS results, and those implicated in other published studies, to provide a more comprehensive study of schizophrenia. Thus, apart from the top hits generated from the gene-based test result in our GWAS, additional candidate SNPs selected from the literature and other schizophrenia research resources were included in our follow-up study. The latter were chosen based on: (i) the top candidate genes as listed in the SZGene Database¹; (ii) previously published GWAS studies and CNV studies for schizophrenia and/or bipolar disorder; (iii) the best hits from the meta-analysis of GWAS studies on schizophrenia and on bipolar disorder as provided by the PGC^{5,15}; and (iv) schizophrenia candidate genes as indicated in plausible pathogenic pathways and our internal research for schizophrenia,

ie, MEK-ERK pathway, AKT signaling pathway and CLOCK pathway. Figure 1 outlines the overall study design. The breakdown of SNP selection by category for the follow-up study is listed in table 1, while more information on SNP selection is provided in the supplementary materials. Genotyping of these SNPs were carried out by Illumina using the GoldenGate Assay. To match the maximum capacity of the GoldenGate assay, a total of 384 SNPs were selected for genotyping.

Quality Control

Similar to the discovery phase, the quality of the SNPs genotyped was assessed by genotyping rates across samples, MAF, and the HWE tests (for control subjects only). We detected 2 SNPs that failed the HWE test (ie, $P < 1 \times 10^{-5}$). All subjects had genotyping rate $> 90\%$. The sex of each subject was confirmed by genotyping data on sex-specific loci provided by Illumina. For duplicated samples, we excluded the one with lower genotyping rate. For family data, Mendelian errors were checked using PLINK.¹⁶ After quality control, we constructed pseudo-controls from the parental genotypes based upon the transmitted and nontransmitted alleles. Thus, in each family genotyped, the offspring affected with schizophrenia was used as a case in the case-control association analysis, while the pseudo-control constructed based on parental genotypes contributed as a control.

Statistical Analysis for Follow-up Study and Meta-analyses

Each successfully genotyped SNP was tested for association by logistic regression analysis. The association test results from the discovery (GWAS) phase and the follow-up study were combined by the inverse variance weighted method (fixed effect). For SNPs selected from previous published GWAS, we also combined the results from the present study (both GWAS and follow-up) with the results given in the previous GWAS. All these procedures were accomplished by PLINK and R. Finally, we calculated the risk allele burden of each individual in the sample as the proportion of risk alleles at all successfully genotyped loci, where the risk alleles were defined from the direction of effect observed in the discovery GWAS or as indicated in previous studies for those selected from previously published literature. A logistic regression test was used to assess the difference in risk allele burden between cases and controls.

Results

Discovery GWAS

After quality control and genotype imputation, the final data set consisted of 2 383 054 SNPs and 2506

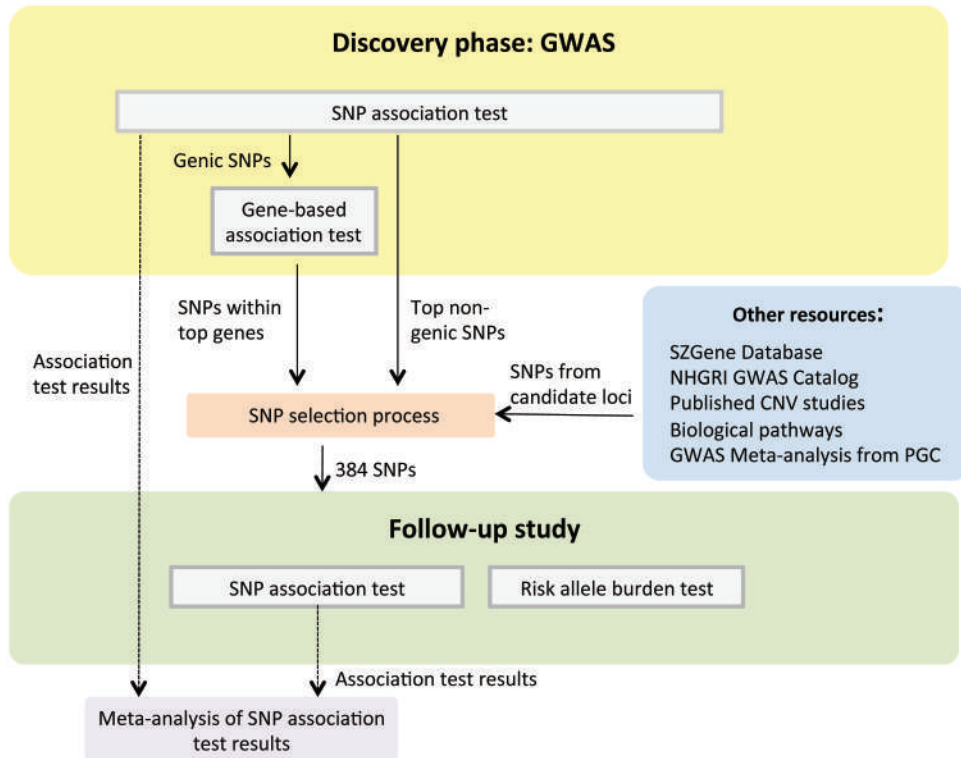


Fig. 1. Outline of our study design. In this project, we adopted a 2-stage study design, the preliminary discovery phase and a follow-up study. In the discovery phase, we performed a case-control genome-wide association study (GWAS) on schizophrenia by carrying out the single nucleotide polymorphism (SNP) and gene-based association tests. Significant genes defined according to the false discovery rate threshold 0.55 were identified. SNPs from the significant genes and top nongenic significant SNPs were selected for genotyping in the follow-up study. Apart from selecting SNP based on our GWAS results, we also included SNPs based on other schizophrenia research findings. In total, we genotyped 384 SNPs in the follow-up study. We then performed association tests on those successfully genotyped SNPs and risk allele burden tests across samples in the follow-up study. Finally, Our GWAS results and follow-up study results were combined as a meta-analysis on schizophrenia.

Table 1. Risk Allele Burden Test Results Per Basis of SNP Selection in the Follow-up Study

Basis of SNP Selection	Number of SNPs Selected for Follow-up Study	Number of SNPs Successfully Genotyped	P values from Risk Allele Burden Tests
Our GWAS on Han Chinese population	130	121	.000458
Other resources			
1. SZGene database	58	53	.706490
2. NHGRI GWAS Catalog and other GWAS studies	81	77	.000146
3. Published CNV studies	10	10	.838960
4. Biological pathways	40	37	.003103
5. GWAS meta-analysis from PGC	65	60	.137151
Total	384	358	

Note: GWAS, Genome-wide association study; SNP, single-nucleotide polymorphism; CNV, copy number variation. Logistic regression tests were performed to compare the risk allele burden of cases and controls for each category of SNP selection. Risk allele burden is defined as the proportion of high-risk alleles present in an individual.

individuals. There were 481 cases (325 males and 156 females) and 2025 controls (674 males and 1351 females). The genomic inflation factor after PCA adjustment (λ) was 1.012, indicating the absence of population stratification. [Supplementary figure](#) shows the Manhattan plot of the GWAS *P* values for all SNPs with MAF ≥ 0.01 . No genome-wide significant association was observed. We

identified 84 SNPs that achieved a *P* value of suggestive significance (ie, $\leq 10^{-5}$), which is more than by random chances (ie, $2\ 383\ 054\ \text{SNPs} \times 10^{-5} = 24\ \text{SNPs}$). For the SNPs on chromosome X, association tests were also performed on females and males separately, and results obtained were similar to those when female and males were analyzed together ([supplementary table](#)).

The gene-based test results are shown in the [supplementary table](#). Among the top 10 candidate genes in our gene-based test results, 3 genes were located within the Xq28 chromosomal region: *RENBP* with a gene-based P value of 5.54×10^{-6} ; *ARHGAP4* with a gene-based P value of 6.24×10^{-5} ; and *MECP2* with a gene-based P value of 6.25×10^{-5} . [Figure 2](#) shows the regional plot of the 500kb stretch within the chromosome Xq28 locus.

Follow-up Study Results

Pseudo-controls were constructed from parents of cases after quality control, The 931 independent controls from Sichuan and 74 from HK, together with the 58 pseudo-controls, formed the control set of 1063 individuals (see [supplementary table](#)). Logistic regression tests were performed on the 358 genotyped SNPs for 1088 patients with schizophrenia (490 males and 598 females) and 1063 controls (489 males, 516 females and 58 pseudo controls), adjusting for the location of sample collection and sex. Pseudo controls were excluded from the association tests on chromosome-X SNPs as they do not have a real sex. We identified 43 SNPs that achieved a 1-tailed P value of nominal significance (ie, ≤ 0.05), which is more than by random chance (ie, the expected number is $358 \text{ SNPs} \times 0.05 = 18 \text{ SNPs}$). Four of them survived the multiple testing correction using 0.1 as the FDR threshold (highlighted in [supplementary table](#)). For the SNPs on chromosome X, association test results obtained were again similar when female and males were analyzed together or separately ([supplementary table](#)).

Meta-analysis Results: Common Variants in Xq28 Confer Risk of Schizophrenia

Log-odds ratios as well as standard errors from our GWAS and the follow-up study were combined using the inverse variance weighted method (fixed-effect). [Table 2](#) lists the 4 significant SNPs revealed from the result of our meta-analysis. As noted, all 4 SNPs are located on the chromosome Xq28 region. The marker rs2269372 (an imputed SNP with an INFO score of 0.97179) reached genome-wide significance (ie, $P < 5 \times 10^{-8}$) with a combined P value of 3.98×10^{-8} and odds ratio (OR) of 1.31 for allele A. This SNP resides in the *RENBP* gene. Three other markers within the same Xq28 region with combined P values close to the genome-wide significance threshold were: rs2269368 ($P_{\text{combined}} = 4.74 \times 10^{-7}$, OR of allele C = 1.25), rs2734647 ($P_{\text{combined}} = 8.78 \times 10^{-7}$, OR of allele C = 1.28), and rs2239464 ($P_{\text{combined}} = 6.71 \times 10^{-7}$, OR of allele G = 1.26). The rs2269368 SNP is located in *ARHGAP4*, while the other 2 SNPs, rs2734647 and rs2239464, are in *MECP2*. The LD between rs2734647 and rs2239464 is strong, with an r^2 of .78 in the follow-up study ([supplementary figure](#)). There was no significant sex difference in allele frequencies of these 4 loci in both the GWAS and follow-up studies as reflected by the heterogeneity test ([supplementary table](#)).

Risk Allele Burden in Follow-up Study

The histogram in [figure 3](#) shows the distributions of risk allele burden in schizophrenia patients and controls. Logistic regression analysis showed that patients with

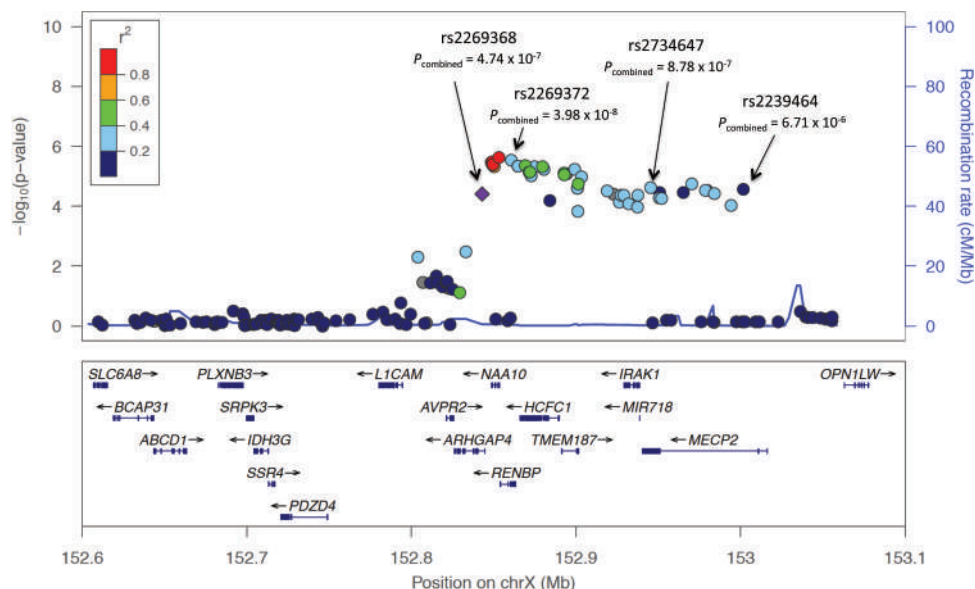


Fig. 2. Regional plot of the associated locus on chromosome Xq28. Association results of our Han Chinese genome-wide association study on schizophrenia, together with recombination rates and genes within the 500kb region on Xq28, were plotted using LocusZoom. The index single nucleotide polymorphism (SNP) rs2269368 is represented by a diamond. The remaining SNPs within this region are labeled with different colors according to their linkage disequilibrium with the index SNP, based on the pair-wise r^2 values from HapMap phaseII JPT + CHB. Combined results of the 4 significant SNPs (ie, rs2269368, rs2269372, rs2734647, and rs2239464) are shown as P_{combined} .

Table 2. Genome-wide Association Study, Follow-up Study, and Meta-analysis Results on the Xq28 Chromosomal Region

Chromosome	Single nucleotide polymorphism	Location (bp)	Risk Allele	Stage 1: Genome-wide association study			Stage 2: Follow-up study			Meta-analysis		Gene
				OR	95% CI*	P Value	OR	95% CI*	P Value (1-tailed)	OR	P Value	
X	rs2269368	152843013	C	1.40	1.21–1.62	3.92×10^{-5}	1.18	1.06–1.32	.0014	1.253	4.74×10^{-7}	<i>ARHGAP4</i>
X	rs2269372	152860739	A	1.62	1.39–1.90	2.84×10^{-6}	1.16	1.02–1.31	.0102	1.313	3.98×10^{-8}	<i>RENBP</i>
X	rs2734647	152945374	C	1.51	1.29–1.78	2.39×10^{-5}	1.15	1.02–1.30	.0107	1.275	8.78×10^{-7}	<i>MECP2</i>
X	rs2239464	153001625	G	1.55	1.31–1.83	2.71×10^{-6}	1.12	0.98–1.27	.0456	1.261	6.71×10^{-6}	<i>MECP2</i>

Note: *95% Confidence interval (CI) of the odds ratio (OR).

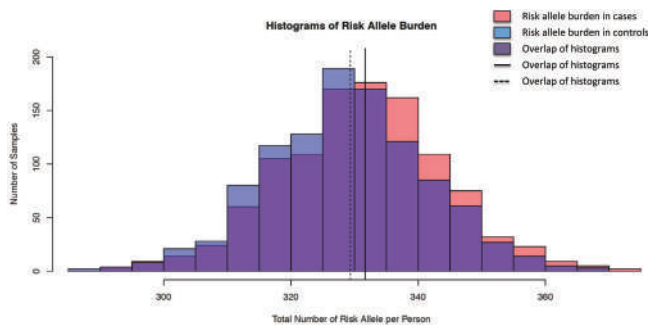


Fig. 3. Histograms showing the risk allele burden of patients with schizophrenia (cases) and controls in the follow-up study. Risk allele burden is defined as the count of risk alleles across 358 successfully genotyped single nucleotide polymorphisms (SNPs) in each individual. Cases harbor significantly more risk alleles in the set of 358 SNPs (mean = 331.7) than controls (mean = 329.3) with a P value of 7.78×10^{-8} in the logistic regression test.

schizophrenia had significantly more risk alleles in the 358 SNPs than controls ($P = 7.78 \times 10^{-6}$). This indicates the existence of schizophrenia susceptibility loci among the SNPs we selected. We also examined the risk allele burdens by category of SNP selection. Results from the risk allele burden test by category (table 1) showed that categories including our GWAS, NHGRI GWAS Catalog, other GWAS studies and genes implicated from plausible biological pathways provided greater insight into the identification of schizophrenia susceptibility loci in the Han Chinese population (with $P = .000146$ – 0.0031).

Comparison Between the Present Study and Previously Published GWASs for Schizophrenia

Some of the SNPs genotyped for our follow-up study were selected from previously published GWAS studies on populations of European ancestry. Unfortunately the 2 GWAS^{4,6} for schizophrenia in the Han Chinese population were published after we had finalized our SNP selection and genotyping.

As a means of consolidation, for the 105 SNPs that were selected based on published GWAS studies, we

combined their P values from the original studies with the P values from the present GWAS and follow-up studies. Supplementary table lists those loci that reached the genome wide significance level after combining the study-based P values. Among which, SNP rs4775413 on chromosome 15, with 1-tailed $P = .00054$ in our follow-up study, had the lowest combined P value of 1.92×10^{-8} . This SNP was previously identified by PGC in their large scale GWAS meta-analysis on schizophrenia.⁵ The genes *RORA* and *VPS13C* are in the closest proximity to this SNP. Two SNPs in *RORA*, rs4774384 and rs11632230, were also genotyped in the follow-up study, and both showed moderate significance with combined P values of 4.419×10^{-6} and 1.02×10^{-3} respectively.

Although none of the susceptibility markers identified in the 2 Han Chinese GWAS had been included in our follow-up study, we have checked their GWAS results against the P values obtained from our GWAS (see supplementary table). We found that the SNPs located on chromosome 1q24.2 and 11p11.2 have the same direction of effect. However, the effect sizes were comparatively smaller in our GWAS indicating a likelihood of population substructure within the Han Chinese population or differential environmental effects. There were little overlaps between the susceptibility loci identified in these 2 Han Chinese GWAS. Recent study on Central Chinese was also not able to replicate the results from these 2 Han Chinese GWAS.¹⁷

The major histocompatibility complex (MHC) has previously been reported to have association with schizophrenia in populations of European ancestry.³ In our GWAS, there were SNPs showing moderate association in the MHC region, but all their P values were $> .0001$ (see supplementary figure). For those SNPs within the MHC region that were also included in our follow-up study, their P values were combined with those reported by previous studies and are listed in supplementary table. It was noted that rs2071287 in the *NOTCH4* gene was the only SNP with nominal significance in the MHC region ($P = .00943$). This was also the only significant SNP in the MHC region identified by a recently published GWAS performed on the Japanese population.¹⁸

The direction of effect with respect to this SNP in our study was consistent with their study.

Discussion

Schizophrenia, as many other psychiatric disorders, is characterized by its multifactorial inheritance. Genetics clearly plays a prominent role in the etiology of schizophrenia as this is indicated both by the high heritability¹⁹ and the higher incidence among the first-degree relatives of affected individuals.^{20,21} Notably, no single genetic locus identified so far could explain the majority of cases demonstrating the complexity and polygenic nature of schizophrenia. Despite the high heritability, epigenetic factors may also exert substantial effects on the development of schizophrenia. In the present study, our focus is placed on the genetic component of schizophrenia.

We have performed a GWAS on a sample of 498 patients with schizophrenia and 2025 controls from the Han Chinese population and a follow-up study using an independent sample of 1027 cases and 1005 controls. Combining the results from both studies, we identified 4 SNPs on chromosome Xq28 with significant association to schizophrenia. Among which, the SNP rs2269372 harbored in *RENBP* (renin binding protein) had achieved genome-wide significance. Two other top hits, rs2734647 and rs2239464 were found collocating in the *MECP2* (methyl CpG binding protein 2 (Rett syndrome)) gene, and both attained high level of significance. The remaining SNP with suggestive significance was rs2269368 which is located in the *ARHGAP4* (Rho GTPase activating protein 4) gene.

RENBP (renin-binding protein) inhibits renin activity by forming a complex with renin. Previous studies have suggested a role of the renin-angiotensin system (RAS) in the pathogenesis of psychiatric disorders including schizophrenia.^{22,23} Angiotensin II is a neurotransmitter that has been shown to interact with dopaminergic systems,^{24,25} and it was postulated that the brain RAS plays a crucial role in normal cognitive functioning.²⁶

Among the 3 candidate genes identified in the Xq28 region, *MECP2* appeared to be the most biologically relevant to schizophrenia. *MECP2* encodes a protein that belongs to the methyl-CpG-binding domain family, which binds to methylated DNA sequences and is able to modify gene expression. It is expressed more abundantly in the brain compared to non-central nervous system (CNS) tissues.²⁷ Mutations in *MECP2* have been associated with several neurodevelopmental disorders²⁸ that include Rett syndrome,^{29,30} autism,³¹ mental retardation,³² neonatal encephalopathy,³³ schizophrenia, and developmental language disorder.³⁴ Joyner et al³⁵ has reported significant sex-specific association of the common rs2239464 marker on *MECP2* with reductions in cortical surface area ($P = .0005$, effect size = -6.17 in males) in subjects with

dementia, affective disorders, psychotic disorder or schizophrenia. According to another recent study by Colibazzi et al,³⁶ reduced surface areas of the cortex was associated with reduced white matter and gray matter in subjects with schizophrenia. Chao et al.³⁷ demonstrated that *MECP2* is crucial for the functioning of GABA-releasing neurons. Mice with *MECP2* deficiency in GABAergic neurons displayed phenotypes similar to Rett syndrome. Notably, it has been hypothesized that GABAergic dysfunction contributes to the pathogenesis of schizophrenia.³⁸⁻⁴¹ A recent large-scale GWAS meta-analysis on schizophrenia has identified a SNP within the *MIR137* (microRNA 137) gene.⁵ Notably, *MECP2* has been found to epigenetically regulate the expression of microRNA miR-137 in adult neural stem cells.⁴² In addition, 4 other loci reaching genome-wide significance in their meta-analysis contained predicted miR-137 target sites.

The other candidate gene on Xq28, *ARHGAP4*, encodes the Rho GTPase activating protein which plays a role in controlling axonal outgrowth and cell motility. This protein is widely expressed in the CNS during embryonic development.⁴³ The chromosomal region Xq28 is an interesting locus that harbors genes which are previously reported to be associated to neurological development or schizophrenia. Another functional candidate in this region is *IRAK1* which encodes interleukin-1 receptor-associated kinase 1, a crucial component of the IL-1 receptor signaling pathway. Interestingly, IL1 beta polymorphisms have been reported to confer risk of schizophrenia.^{44,45} Previous studies also suggested that IL-1beta may be involved in the response to prenatal infection and hence affecting neurodevelopment.⁴⁶

The present study adopted a 2-stage approach, where the follow-up study included SNPs from other existing research projects in addition to the findings from our GWAS. By using the data in the follow-up study, we have shown that patients with schizophrenia have significantly higher number of risk alleles than controls with respect to the SNPs that we have selected. This indicated that at least a proportion of the putative risk alleles are highly likely to be associated with schizophrenia. This finding lends further support to the polygenic basis for schizophrenia. Byun et al demonstrated that subjects with genetic high risk for schizophrenia have cortical thinning in the right anterior cingulate cortex and left paracingulate cortex, which may represent neurodevelopmental alterations that result from genetic liability for schizophrenia.⁴⁷

Our discovery GWAS was limited by the low statistical power due to small sample size. In light of this, we attempted to improve the power by employing the gene-based analysis. In addition, evidence for schizophrenia susceptibility loci and genes from other research resources were also considered and incorporated in our follow-up study. This enabled us to have a more comprehensive study of the genetics of schizophrenia. The little overlap

between the 2 published GWAS studies on the Han Chinese^{48,49} and the unsuccessful replication of the signal reported by Ma et al¹⁷ suggested that there is an urge of performing meta-analysis of GWASs on the Han Chinese population in order to gain higher statistical power.

By combining our results with the published findings, we also identified an autosomal candidate SNP, rs4775413, on chromosome 15 which reached high level of significance. This SNP is located in the upstream region of *RORA* and was suggested to be associated with depression and bipolar disorder.^{50,51} Two other SNPs (rs4774384 and rs11632230) in the *RORA* genes also showed moderate significance in the combined analysis of our GWAS and follow-up study. It is also interesting to note that both *RORA* and *MECP2* belong to the clock pathway which was suggested to be associated with depression.⁵²

Our follow-up study was characterized by the incorporation of evidence for schizophrenia from other external research resources, and it gave some insights into knowledge-based SNP prioritization. Based on the risk allele burden tests, we found that SNPs selected from our GWAS, NHGRI GWAS Catalog, other GWAS studies and genes implicated from plausible biological pathways in general contribute more to our schizophrenia study. On the other hand, this may reflect possible bias in population background for candidate SNPs that were reported in SZGene database and PGC meta-analysis.

So far there is no golden standard established yet in analyzing the genotypes on chromosome X, though 2 approaches have been proposed: (a) combined analysis: analysis performed on male and female genotypes together; (b) stratified analysis: analyses performed on male and female genotypes separately, and the results are combined by using the meta-analysis method of inverse weighted variance. Regarding the combined analysis, male genotypes can be coded as 0, 1 or 0, 2. The male genotype coding 0, 2 is based on the assumption of random X-chromosome inactivation in females, such that only one allele is expressed from each locus. The allele to be expressed is selected randomly during early fetal development. As a result, the effect of an allele in females is essentially halved (or females only have one active X chromosome, effectively, but we do not know which one). To compare the association test results generated from these different approaches, we performed a simulation which showed that results from stratified analysis with different male genotype codings for SNPs on chromosome X were similar. Apart from the combined analysis, we also performed the stratified analyses on males and females in our 2-stage study, and combined the results using the inverse variance weighted method. Association test results obtained were similar when females and males were analyzed together or separately.

In summary, we have identified a novel susceptibility locus for schizophrenia on Xq28 in the Han Chinese population. In addition, the risk allele burden test provides further support for a polygenic basis for schizophrenia.

Supplementary Material

Supplementary material is available at <http://schizophreniabulletin.oxfordjournals.org>.

Funding

Hong King Research Grants Council General Research Fund (774707M, 777511M), The University of Hong Kong Small Project Funding (201007176248, 201007176166), The University of Hong Kong Strategic Research Fund on Genomics, European Community Seventh Framework Programme Grant on European Network of National Schizophrenia Networks Studying Gene-Environment Interactions (EU-GEI), Croucher Foundation Scholarship (to H.C.S), National Nature Science Foundation of China (81130024, 30530300, and 30125014, TL), National Key Technology R & D Program of the Ministry of Science and Technology of China during the 12th Five-Year Plan (2012BAI01B06, to T.L.), the Ph.D. Programs Foundation of Ministry of Education of China (20110181110014, to T.L.).

Acknowledgment

The authors declare no conflict of interest.

References

- Allen NC, Bagade S, McQueen MB, et al. Systematic meta-analyses and field synopsis of genetic association studies in schizophrenia: the SzGene database. *Nat Genet*. 2008;40:827–834.
- Kirov G, Zaharieva I, Georgieva L, et al. A genome-wide association study in 574 schizophrenia trios using DNA pooling. *Mol Psychiatry*. 2009;14:796–803.
- Purcell SM, Wray NR, Stone JL, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*. 2009;460:748–752.
- Shi Y, Li Z, Xu Q, et al. Common variants on 8p12 and 1q24.2 confer risk of schizophrenia. *Nat Genet*. 2011;43:1224–1227.
- Schizophrenia Psychiatric Genome-Wide Association Study Consortium. Genome-wide association study identifies five new schizophrenia loci. *Nat Genet*. 2011;43:969–976.
- Yue WH, Wang HF, Sun LD, et al. Genome-wide association study identifies a susceptibility locus for schizophrenia in Han Chinese at 11p11.2. *Nat Genet*. 2011;43:1228–1231.
- Xiao SM, Kung AW, Gao Y, et al. Post-genome wide association studies and functional analyses identify association of MPP7 gene variants with site-specific bone mineral density. *Hum Mol Genet*. 2012;21:1648–1657.
- Guo Y, Tomlinson B, Chu T, et al. A genome-wide linkage and association scan reveals novel loci for hypertension and blood pressure traits. *PLoS One*. 2012;7:e31489.
- Chan KY, Wong CM, Kwan JS, et al. Genome-wide association study of hepatocellular carcinoma in Southern Chinese patients with chronic hepatitis B virus infection. *PLoS One*. 2011;6:e28798.
- Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for

- stratification in genome-wide association studies. *Nat Genet.* 2006;38:904–909.
11. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* 2009;5:e1000529.
 12. Clayton DG. Sex chromosomes and genetic association studies. *Genome Med.* 2009;1:110.
 13. Clayton D. Testing for association on the X chromosome. *Biostatistics.* 2008;9:593–600.
 14. Li MX, Gui HS, Kwan JS, Sham PC. GATES: a rapid and powerful gene-based association test using extended Simes procedure. *Am J Hum Genet.* 2011;88:283–293.
 15. Psychiatric GWAS Consortium Bipolar Disorder Working Group. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nat Genet.* 2011;43:977–983.
 16. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81:559–575.
 17. Ma L, Tang J, Wang D, et al. Evaluating risk loci for schizophrenia distilled from genome-wide association studies in Han Chinese from central China. *Mol Psychiatry.* 2011;18:638–639.
 18. Ikeda M, Aleksic B, Yamada K, et al. Genetic evidence for association between NOTCH4 and schizophrenia supported by a GWAS follow-up study in a Japanese population. *Mol Psychiatry.* 2013;18:636–638.
 19. Sullivan PF, Kendler KS, Neale MC. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Arch Gen Psychiatry.* 2003;60:1187–1192.
 20. McGuffin P, Owen MJ, Farmer AE. Genetic basis of schizophrenia. *Lancet.* 1995;346:678–682.
 21. Lichtenstein P, Yip BH, Björk C, et al. Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. *Lancet.* 2009;373:234–239.
 22. Crescenti A, Gassó P, Mas S, et al. Insertion/deletion polymorphism of the angiotensin-converting enzyme gene is associated with schizophrenia in a Spanish population. *Psychiatry Res.* 2009;165:175–180.
 23. Kucukali CI, Aydin M, Ozkok E, et al. Angiotensin-converting enzyme polymorphism in schizophrenia, bipolar disorders, and their first-degree relatives. *Psychiatr Genet.* 2010;20:14–19.
 24. Jenkins TA, Allen AM, Chai SY, MacGregor DP, Paxinos G, Mendelsohn FA. Interactions of angiotensin II with central dopamine. *Adv Exp Med Biol.* 1996;396:93–103.
 25. Jenkins TA, Mendelsohn FA, Chai SY. Angiotensin-converting enzyme modulates dopamine turnover in the striatum. *J Neurochem.* 1997;68:1304–1311.
 26. Wright JW, Harding JW. The brain angiotensin system and extracellular matrix molecules in neural plasticity, learning, and memory. *Prog Neurobiol.* 2004;72:263–293.
 27. LaSalle JM, Goldstine J, Balmer D, Greco CM. Quantitative localization of heterogeneous methyl-CpG-binding protein 2 (MeCP2) expression phenotypes in normal and Rett syndrome brain by laser scanning cytometry. *Hum Mol Genet.* 2001;10:1729–1740.
 28. Chahrour M, Jung SY, Shaw C, et al. MeCP2, a key contributor to neurological disease, activates and represses transcription. *Science.* 2008;320:1224–1229.
 29. Amir RE, Van den Veyver IB, Wan M, Tran CQ, Francke U, Zoghbi HY. Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nat Genet.* 1999;23:185–188.
 30. Williamson SL, Christodoulou J. Rett syndrome: new clinical and molecular insights. *Eur J Hum Genet.* 2006;14:896–903.
 31. Shibayama A, Cook EH Jr, Feng J, et al. MECP2 structural and 3'-UTR variants in schizophrenia, autism and other psychiatric diseases: a possible association with autism. *Am J Med Genet B Neuropsychiatr Genet.* 2004;128B:50–53.
 32. Gécz J, Shoubridge C, Corbett M. The genetic landscape of intellectual disability arising from chromosome X. *Trends Genet.* 2009;25:308–316.
 33. Geerdink N, Rotteveel JJ, Lammens M, et al. MECP2 mutation in a boy with severe neonatal encephalopathy: clinical, neuropathological and molecular findings. *Neuropediatrics.* 2002;33:33–36.
 34. Cohen D, Lazar G, Couvert P, et al. MECP2 mutation in a boy with language disorder and schizophrenia. *Am J Psychiatry.* 2002;159:148–149.
 35. Joyner AH, J CR, Bloss CS, et al. A common MECP2 haplotype associates with reduced cortical surface area in humans in two independent populations. *Proc Natl Acad Sci U S A.* 2009;106:15483–15488.
 36. Colibazzi T, Wexler BE, Bansal R, et al. Anatomical abnormalities in gray and white matter of the cortical surface in persons with schizophrenia. *PLoS One.* 2013;8:e55783.
 37. Chao HT, Chen H, Samaco RC, Xue M, Chahrour M, Yoo J et al. Dysfunction in GABA signalling mediates autism-like stereotypies and Rett syndrome phenotypes. *Nature* 2010;468:263–269.
 38. Guidotti A, Auta J, Davis JM, Dong E, Grayson DR, Veldic M, et al. GABAergic dysfunction in schizophrenia: new treatment strategies on the horizon. *Psychopharmacology (Berl)* 2005;180:191–205.
 39. Blum BP, Mann JJ. The GABAergic system in schizophrenia. *The Int J Neuropsychopharmacol official scientific journal of the Collegium Internationale Neuropsychopharmacologicum (CINP)* 2002;5:159–179.
 40. Costa E, Davis JM, Dong E, Grayson DR, Guidotti A, Tremolizzo L, et al. A GABAergic cortical deficit dominates schizophrenia pathophysiology. *Crit Rev Neurobiol* 2004;16:1–23.
 41. Nakazawa K, Zsiros V, Jiang Z, Nakao K, Kolata S, Zhang S, et al. GABAergic interneuron origin of schizophrenia pathophysiology [published online ahead of print 2011]. *Neuropharmacology.*
 42. Szulwach KE, Li X, Smrt RD, et al. Cross talk between microRNA and epigenetic regulation in adult neurogenesis. *J Cell Biol.* 2010;189:127–141.
 43. Vogt DL, Gray CD, Young WS 3rd, Orellana SA, Malouf AT. ARHGAP4 is a novel RhoGAP that mediates inhibition of cell motility and axon outgrowth. *Mol Cell Neurosci.* 2007;36:332–342.
 44. Sasayama D, Hori H, Teraishi T, et al. Possible association between interleukin-1 β gene and schizophrenia in a Japanese population. *Behav Brain Funct.* 2011;7:35.
 45. Söderlund J, Schröder J, Nordin C, et al. Activation of brain interleukin-1 β in schizophrenia. *Mol Psychiatry.* 2009;14:1069–1071.
 46. Marx CE, Jarskog LF, Lauder JM, Lieberman JA, Gilmore JH. Cytokine effects on cortical neuron MAP-2 immunoreactivity: implications for schizophrenia. *Biol Psychiatry.* 2001;50:743–749.

47. Byun MS, Kim JS, Jung WH, et al. Regional cortical thinning in subjects with high genetic loading for schizophrenia. *Schizophr Res.* 2012;141:197–203.
48. Shi Y, Li Z, Xu Q, et al. Common variants on 8p12 and 1q24.2 confer risk of schizophrenia. *Nat Genet.* 2011;43:1224–1227.
49. Yue WH, Wang HF, Sun LD, et al. Genome-wide association study identifies a susceptibility locus for schizophrenia in Han Chinese at 11p11.2. *Nat Genet.* 2011;43:1228–1231.
50. Le-Niculescu H, Patel SD, Bhat M, et al. Convergent functional genomics of genome-wide association data for bipolar disorder: comprehensive identification of candidate genes, pathways and mechanisms. *Am J Med Genet B Neuropsychiatr Genet.* 2009;150B:155–181.
51. Terracciano A, Tanaka T, Sutin AR, et al. Genome-wide association scan of trait depression. *Biol Psychiatry.* 2010;68:811–817.
52. Soria V, Martínez-Amorós E, Escaramís G, et al. Differential association of circadian genes with mood disorders: CRY1 and NPAS2 are associated with unipolar major depression and CLOCK and VIP with bipolar disorder. *Neuropsychopharmacology.* 2010;35:1279–1289.