

Recent Positive Selection Drives the Expansion of a Schizophrenia-associated Variant within 10q24.33 in Human Populations through Its Pleiotropic Effects on Diverse Human Complex Traits

Lingyi Li¹, Hong Chang¹, Tao Huang^{2,3}, Yan Zheng², Lu Qi^{2,4}, Xiao Xiao¹, Ming Li^{1*}

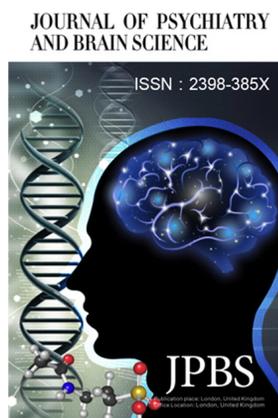
¹ Key Laboratory of Animal Models and Human Disease Mechanisms of the Chinese Academy of Sciences and Yunnan Province, Kunming Institute of Zoology, Kunming, Yunnan, China;

² The Department of Nutrition, Harvard School of Public Health, Boston, MA, USA;

³ Epidemiology Domain, Saw Swee Hock School of Public Health and Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore;

⁴ Department of Epidemiology, School of Public Health and Tropical Medicine, Tulane University, New Orleans, LA, USA.

*Corresponding Author: Ming Li, Ph.D., Kunming Institute of Zoology, Chinese Academy of Sciences, 32 East Jiao-Chang Rd., Kunming, Yunnan 650223, China. Email: limingkiz@mail.kiz.ac.cn.



<http://jpbs.qingres.com>

 OPEN ACCESS

DOI: 10.20900/jpbs.20170001

Received: November 15, 2016

Accepted: January 12, 2017

Published: February 25, 2017

Copyright: ©2017 Cain *et al.* This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

Background: Genome-wide scans have revealed that Darwinian natural selection affects human genome during human evolution, helping us to optimize complex human adaptations. Natural selection is also reflected in significant associations with human phenotypes, e.g., single genomic locus associated with diverse complex disorders and traits.

Method: Meta-analyses of single nucleotide polymorphisms (SNPs) in 10q24.33 with schizophrenia in Asian populations were conducted. Population genetic analyses of the schizophrenia risk SNPs were also conducted in diverse populations.

Results: We report a SNP rs11191580 in 10q24.33 to be significantly

associated with schizophrenia in Asians. Further pleiotropic analyses reveal that the schizophrenia non-risk allele [C] at rs11191580 also reduces risk of blood pressure and hypertension in both European and Asian populations. The non-risk allele [C] is present in populations from Europe and Asia mainland, but totally absent in Africans. Further evolutionary analyses show that the derived allele (i.e., C-allele) has experienced recent positive selection in Europeans and Asians. Through subsequent exploratory analysis, we propose that the colder environment in Europe and Asia is likely the selection pressure, i.e., when modern humans migrated “out of Africa” and moved to Europe and Asia mainland (the continent colder than Africa), new allele derived due to positive selection. This new allele probably helped human beings to adapt to the cold environment and also protected them from risk of hypertension.

Conclusion: The current data provides an intriguing example demonstrating the reflection of natural selection on pleiotropic genomic locus, illustrating a possible mechanism of selection pressures in human populations.

KEY WORDS: Schizophrenia, hypertension, obesity, 10q24.33, positive selection

1 INTRODUCTION

Schizophrenia is a severe and common complex disorder with a lifetime prevalence of ~ 0.5 % among world populations.^[1] Though heritability estimate of schizophrenia exceeds 80 %, ^[2] alleles identified to date explain only a small proportion of this rate. Until recently, large-scale genome-wide association studies (GWAS) have explained the polygenic architecture of schizophrenia,^[3] i.e. a large number of single nucleotide polymorphisms (SNPs) having low to moderate effect sizes and are variable in statistical associations in small sample set altogether account for the disease risk.

In addition to the increasing number of reported schizophrenia associated SNPs, epidemiological studies have also identified multiple traits and diseases strongly associated with this illness, including blood pressure, hypertension and cardiovascular diseases etc.^[4,5] The etiologic relationship between these phenotypes is yet to be explained. One plausible hypothesis is the genetic pleiotropy,^[6] which is the association of individual genetic variation with ≥ 2 phenotypes. In fact, there

are often overlapping traits among the above-mentioned phenotypes defined through behavioral or clinical investigations, indicating potential shared genetic influences.

However, the underlying evolutionary mechanisms of these genetic pleiotropic phenomena remain unclear. Darwinian natural selection has played (or is still playing) important roles in complex human adaptations during evolution, resulting in presence or disappearance of specific allele that could help us adapt to external environment or survive through severe disorders. Among these genomic loci affected by natural selection, many of them can still be identified through association studies with human phenotypes such as microcephaly.^[7,8] Also, loci showing pleiotropic effects are recognized as potential signatures that Darwinian natural selection might have acted on this genomic region.^[9,10] For example, we previously reported genetic pleiotropic effects of a schizophrenia SNP rs13107325 in SLC39A8 on diverse human diseases and traits including schizophrenia, obesity and blood pressure, which were likely resulted from the concomitantly observed strong Darwinian positive selection on this genomic locus.^[11]

Chromosome 10q24.33 is one of genomic regions most significantly associated with schizophrenia in recent GWAS among European populations,^[3] but the associations in other ethnic groups are still unclear. Here, we have performed a series of convergent analyses on the evolutionary mechanisms underlying this genomic region. First, using a large-scale analysis in nine independent samples, we confirmed that the risk SNP rs11191580 in 10q24.33 was also significantly associated with schizophrenia in Asian samples, supporting the shared genetic risk of schizophrenia on this locus between continental populations. We then assessed the impact of the risk SNP rs11191580 on local gene expression in brain samples, and found that the SNP was associated with the expression of a human-specific isoform of AS3MT. We also showed that rs11191580 in 10q24.33 region exhibited pleiotropic effects on diverse complex traits and diseases in both European and Asian populations, i.e., the schizophrenia non-risk allele [C] of rs11191580 was associated with decreased risk of blood pressure, hypertension and cardiovascular diseases. In addition, we demonstrated that the schizophrenia non-risk allele [C] of rs11191580 was present in both Europeans and Asians but absent in Africans, and the C-allele underwent recent positive selection in the populations from Europe and Asia mainland. This result indicated an origination of the allele after the modern human’s “out-of-Africa” scenario, and

suggested that the selection pressure might be related to the local environmental factors in Europe and Asia, e.g., cold temperature of the local climate. This study further confirmed the theory that human evolutionary history acts on the changes of allele frequency of specific genomic loci and results in pleiotropic effects in complex human disorders/traits.

2 MATERIALS AND METHODS

2.1 Schizophrenia Samples

For the genetic analysis of rs11191580 with schizophrenia in Asian populations, seven case-control and two family-based samples were used, including a total of 13,042 cases, 18,435 controls and 1,115 nuclear families. All patients were diagnosed with schizophrenia according to either International Classification of Diseases 10 (ICD-10) or Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) criteria, and control subjects were local volunteers with no record of mental illnesses.

Shaanxi, China. The Shaanxi sample included 1,430 schizophrenia patients and 1,570 controls. The subjects were Han Chinese recruited from First Affiliated Hospital of Xi'an Jiaotong University and Xi'an Mental Health Center. 53.4 % of the patients, were males (mean age 37.6 ± 11.2 years) and 46.6 % were females (mean age 35.1 ± 10.6 years). All patients were diagnosed by at least three psychiatrists according to the DSM-IV criterion. The control sample consisted of 54.8 % males (mean age 32.9 ± 10.9 years) and 45.2 % females (mean age 34.2 ± 11.7 years). Rs11191580 was genotyped with matrix-assisted laser desorption ionization–time-of-flight method by Guan *et al.*^[12]

Yunnan, China. The Yunnan sample consisted of 547 unrelated schizophrenia patients (mean age 38.7 ± 10.6 years) and 755 unrelated healthy control subjects (mean age 36.8 ± 6.7 years). Rs11191580 was previously genotyped with SNaPShot method in this sample.^[13] The patients were recruited from The Second People's Hospital of Yuxi City and diagnosed as having schizophrenia by at least two psychiatrists according to the ICD-10 criterion. Potential participants who had the history of alcoholism, epilepsy, neurological diseases, or drug abuse were excluded. Meanwhile, unrelated healthy subjects were recruited from local community as control samples. These controls were all free of psychiatric disorders, drug abuse, alcohol dependence, or brain injury. All the patient and control subjects were Han Chinese.

Hunan, China. The Hunan sample contained a total of 976 unrelated schizophrenia patients (mean age 24.9 ± 8.3 years) and 1,043 matched healthy controls (mean age 37.4 ± 14.2 years). All subjects were Han Chinese recruited from Hunan Province in South Central China. In brief, patients were diagnosed independently by two experienced psychiatrists as schizophrenia according to DSM-IV criteria. Rs11191580 was genotyped with SNaPShot method in this sample by Ma *et al.*^[14]

BIOX, China. The BIOX sample was comprised of two independent cohorts of 7,335 cases and 11,964 controls in total.^[15] Participants in the schizophrenia group were from various mental health centers. The patients were interviewed by two independent psychiatrists and were diagnosed according to DSM-IV criteria. All healthy controls were randomly selected from Han Chinese volunteers who were asked to reply to a written invitation to evaluate their medical histories. The rs11191580 was not included in the original BIOX study, but a proxy SNP (rs732998, $r^2 = 1.0$ with rs11191580 in Chinese) was genotyped, with Affymetrix Genome-Wide Human SNP Array and MassARRAY iPLEX Gold platform respectively.

Sichuan, China. The Sichuan sample was comprised of 1,087 schizophrenia cases and 1,063 healthy controls. All subjects recruited were of Han Chinese origin. Patients with schizophrenia were recruited from West China Hospital of Sichuan University in Sichuan, China and were assessed using the Structured Clinical Interview for DSM-IV Patient Version (SCID-P). Included patients fulfilled the diagnostic criteria for schizophrenia as specified in DSM-IV. The rs11191580 was genotyped with Illumina GoldenGate Assay in this sample.^[16]

Guangxi, China. The Guangxi sample contained 462 schizophrenia patients and 598 healthy controls of Han Chinese origin as described previously.^[17] The diagnosis of SCZ was confirmed independently by two experienced psychiatrists following the DSM-IV criteria. Rs11191580 was genotyped using Sequenom Mass-ARRAY® technology (Sequenom Inc., San Diego, CA, USA).

Asia family-based sample 01. This sample consisted of 536 affected families, comprising of 1,633 members including 698 schizophrenia patients. Rs11191580 was genotyped with TaqMan assay in this sample by Takahashi *et al.*^[18] Diagnoses of all affected family members were based on either ICD-10 or DSM-IV criteria.

Asia family-based sample 02. This sample was comprised of 579 affected families of 2,296 individuals. Rs11191580 was genotyped in this

sample using Illumina iSelect assay by Aberg *et al.*^[19] Diagnoses of all affected family members were based on DSM-III-R diagnosis criteria.

Aichi, Japan. The Aichi sample included 1,032 unrelated schizophrenia patients (mean age 46.8 ± 14.8 years) and 993 unrelated healthy control subjects (mean age 49.7 ± 14.0 years). Rs11191580 was genotyped with Sequenom iPLEX Gold system (Sequenom, San Diego, CA) in this sample by Saito *et al.*^[20] All the patients and control subjects were of Japanese origin. Patients were diagnosed as having schizophrenia according to the DSM-IV criteria.

Osaka, Japan. The Osaka sample was comprised of 173 schizophrenia patients (mean age 36.0 ± 12.3 years) and 449 healthy controls (mean age 35.4 ± 12.8 years). Rs11191580 was genotyped using TaqMan assay by Ohi *et al.*^[21] All subjects were biologically unrelated within the second-degree of relationship and were of Japanese descent. Each patient with schizophrenia was diagnosed by at least two psychiatrists according to the criteria from DSM-IV.

2.2 Statistical Analysis

We used odds ratio (OR) and standard error (SE) from each individual cohort to estimate heterogeneity between different samples and to calculate the pooled OR and 95 % confidence interval (CI) in the combined samples.

In case-control studies, the allelic-specific OR and its associated SE of rs11191580 were calculated according to the number of allele accounts (or extracted from the studies) using RevMan 5.0.

In family-based studies, the OR of transmission disequilibrium test (TDT) and its associated SE were calculated as follows:^[22]

$$OR=a/b$$

$$SE=\text{SQRT}(1/a+1/b)$$

Where “*a*” means number of times that A1 allele is transmitted from heterozygous parents to affected proband, and “*b*” means number of times that A1 allele is not transmitted.

To combine the results of individual samples, we calculated the heterogeneity between each study using the Cochran’s (Q) χ^2 test, which is a weighted sum of the squares of the deviations of individual OR estimates from the overall estimate. In the absence of heterogeneity among individual studies, we used a fixed-effect model to combine the samples and to calculate the pooled OR and corresponding 95 % CIs; otherwise, a random-effect model was applied.

The meta-analysis was performed with the classical Inverse-Variance weighted methods using the R package (Meta module). We used a forest plot to graphically present the pooled ORs and the 95 % CIs. Each study was represented by a square in the plot, and the weight of each study was also shown.

2.3 MRI Imaging Analysis

We recruited 294 unrelated healthy Han Chinese subjects (187 females and 107 males, mean age 35.1 ± 12.5 years). All volunteers were free of any history of mental disorders, drug abuse, alcohol dependence or brain injury. All individuals have provided written informed consents for participation. This imaging sample was previously shown to be effective for the detection of genetic effects on the features extracted from magnetic resonance imaging (MRI) data with negligible population stratification.^[13,23] Structural MRI data were acquired using a Philips MRI scanner (Achieva Release 3.2.1.0) operating at 3 Tesla. High-resolution whole brain volumetric T1-weighted images were acquired sagittally with an inversion-recovery prepared 3-D spoiled gradient echo (SPGR) pulse sequence (Repetition Time = 7.38 ms, Echo Time = 3.42 ms, flip angle = 8, voxel dimensions = $1.04 \times 1.04 \times 1.80$ mm³, slice thickness = 1.2 mm).

We used FSL-FIRST (FMRIB’s Integrated Registration and Segmentation Tool) from FMRIB Software Library (FSL) package to calculate the brain volume.^[24] The shape/appearance models used in FIRST were constructed for 15 different subcortical structures using 336 manually segmented images. During registration, the input data (3D T1 images) were transformed to the MNI (Montreal Neurological Institute) 152 standard space by means of affine transformations based on 12 degrees of freedom. After subcortical registration, a sub-cortical mask was applied to locate the different subcortical structure. Segmentation based on shape models and voxel intensities was then conducted. All segmented subcortical regions were examined visually for registration or segmentation errors. Intracranial volume (ICV) as well as gray matter and white matter volume were calculated by removing non-brain material and segment into gray matter/white matter/CSF using FSL-FAST.^[25] Multiple linear regression analysis was performed to compare the brain volumes among different genotypes of rs11191580. Sex, age, age², sex×age, sex×age², and four principal components were included as covariates.

2.4 Expression Quantitative Trait Loci (eQTL)

Analysis

For eQTL analysis, we used the gene expression data from a recently published study^[26] of 135 postnatal Caucasian non-psychiatric controls. In brief, the postmortem dorsolateral prefrontal cortex (DLPFC) samples were dissected and homogenized. 80–120M 100 bp paired end reads RNA-sequencing was then performed using Illumina Hi-Seq 2000 according to manufacturer's protocol. Paired-end reads of cDNA sequences were aligned to the human genome reference (UCSC hg19) by splice-read mapper (TopHat v2.0.4),^[27] providing known transcripts from Ensembl Build GRCh37.67. To quantify mRNA expression, we considered only the alignments in the 10q24.33 genomic region (corresponding to chr10:103605356–105615301 on genome build hg19). We controlled for expression heterogeneity due to potential latent factors using principal component analysis (PCA). Principal components (PCs) were calculated using the transformed (log base 2 with an offset of 1) read counts in chr10:103605356–105615301 region, and the first 10 PCs were used as covariates in the eQTL analyses. Genomic principal components were also calculated based on the genomic variants across common linkage disequilibrium independent variants across the genome, and the first five PCs were used as covariates in the eQTL analyses. Statistical analyses of mRNA expression associated with genotypes were conducted in R 3.0.1 using linear regression, treating log₂-transformed expression levels as the outcome, and covaried for 10 expression PCs and five genomic PCs.

2.5 Analyses of Pleiotropic Effects of rs11191580

2.5.1 Coronary Artery Disease

In 2012, the CARDIoGRAMplusC4D Consortium performed an association analysis in 63,746 coronary artery disease (CAD) cases and 130,681 controls of European or Asian descent,^[28] and identified 15 loci reaching genome-wide significance. The total number of susceptibility loci for CAD returned by this analysis was 46, and a further 104 independent variants were found to be strongly associated with CAD at a 5 % false discovery rate (FDR).

2.5.2 Blood Pressure and Hypertension

The International Consortium for Blood Pressure

(ICBP) performed a multi-stage GWAS meta-analysis of blood pressure and hypertension in > 20,000 individuals of European ancestry. Phenotypes included in the analysis were systolic blood pressure (SBP), diastolic blood pressure (DBP), pulse pressure (PP, the difference between SBP and DBP), mean arterial pressure (MAP, a weighted average of SBP and DBP) and hypertension.^[29,30]

In another study, Kato *et al.*^[31] conducted a meta-analysis of GWAS of SBP and DBP in >19,000 subjects of east Asian ancestry from the AGEN-BP consortium, followed by independent replications in east Asian samples.

2.5.3 Body Mass Index and Obesity

Obesity is a highly prevalent and heritable trait in global populations.^[32] As a non-invasive and inexpensive measure of obesity, body mass index (BMI) has been extensively used to predict the risk of obesity-related complications. To better understand the biological basis of obesity, Speliotes *et al.*^[33] conducted genetic association analyses between BMI and about 2.8 million SNPs in up to 123,865 individuals of European ancestry. They followed up 42 SNPs in up to 125,931 additional subjects and identified 18 new loci showing genome-wide significant association with obesity. Detailed information about sample ascertainment and statistical analyses can be found in the original study^[33] and GIANT website (http://www.broadinstitute.org/collaboration/giant/index.php/Main_Page).

Meanwhile, Wen *et al.*^[34] conducted a meta-analysis of associations between BMI and ~ 2.5 million SNPs among 86,757 individuals of Asian ancestry, followed by *in silico* and *de novo* replication among 7,488–47,352 additional Asian-ancestry individuals, by which they identified several novel BMI-associated loci in Asians.

2.5.4 Energy Intake

Dietary intake of macronutrients (carbohydrate, protein, and fat) is associated with risk of chronic conditions such as obesity and diabetes.^[35] Two GWASs (discovery sample size $N > 30\,000$ in each study) were conducted in independent samples of European descent to identify common genetic variants associated with macronutrients intake.^[36,37] The authors reported variants in FGF21 and FTO to exhibit genome-wide significant associations.

2.6 Samples for Population Genetic Analyses

We used a total of 1,092 unrelated individuals from the 1000-Human-Genome Project,^[38] including 379 Europeans (including 85 Utah residents with Northern and Western European ancestry; 88 Toscana from Italia; 14 Iberian from Spain; 93 Finnish from Finland; and 89 from England and Scotland), 286 Asians (consisted of 97 Han Chinese from Beijing, China; 100 Southern Han Chinese; and 89 Japanese from Tokyo, Japan), 181 Americans (comprised of 66 of Mexican ancestry from Los Angeles USA; 55 Puerto Ricans from Puerto Rico; and 60 Colombians from Medellin, Colombia) and 246 Africans (containing 88 Yoruba from Ibadan, Nigeria; 97 Luhya from Webuye, Kenya; 61 of African ancestry from Southwest USA), to analyze the frequency distribution of rs11191580 in different geographic populations, and to calculate the integrated haplotype score (iHS).

2.7 Tests of Natural Selection

The iHS is a test typically designed to capture incomplete selective sweep for detection of recent positive selection. The iHS test statistic, which is calculated for common SNPs (minor allele frequency > 0.05) in the candidate region, reflects the differences in the long-range LD patterns containing the ancestral versus derived alleles. The genotype data was first phased, and then the iHS was calculated using software Selscan according to the methods implemented in previous studies.^[39-42] SNPs under positive selection should have $|iHS| > 2$ ($iHS < -2$ means the derived allele undergoes positive selection and $iHS > 2$ indicates the ancestral allele undergoes positive selection), which corresponded to the most extreme 5 % of iHS values across the genome with minor allele frequency > 0.05.

3 RESULTS

3.1 Rs11191580 was Significantly Associated with Schizophrenia and White Matter Volume in Asian Populations

In 2011, rs11191580 was reported to be significantly associated with schizophrenia in the GWAS by

Psychiatric Genomics Consortium (PGC1) in populations of European descent ($p = 1.11 \times 10^{-8}$, OR = 0.870),^[43] and the significance became stronger in the later European GWAS meta-analysis (PGC1 + SWE) with an increased sample size ($p = 2.09 \times 10^{-10}$, OR = 0.827).^[44] However, such associations were analyzed in Europeans, and the situations in other ethnic groups were rather limited with only one replication study in a smaller Han Chinese sample by a research group from Shaanxi, China ($p = 3.03 \times 10^{-4}$, OR = 0.808).^[12]

Here, to further confirm if rs11191580 was associated with schizophrenia in general and larger Asian populations, we used additional data of rs11191580 from nine independent Asian samples with no overlap with the Shaanxi sample. Power analysis indicated that the present sample size showed a > 80% power of detecting significant association ($p < 0.05$). In these combined samples including 11,612 cases, 16,865 controls and 1,115 nuclear families, rs11191580 was significantly associated with schizophrenia ($p = 1.99 \times 10^{-3}$, OR = 0.942, 95%CI = 0.907–0.978). When the Shaanxi sample was included in the meta-analysis, the association became stronger in a total of 13,042 cases, 18,435 controls and 1,115 families ($p = 4.78 \times 10^{-5}$, OR = 0.928, 95%CI = 0.895 – 0.962). There was no heterogeneity among these samples ($I^2 = 16.1\%$, $p = 0.295$), and sensitivity analyses did not find changes of the significant association after removal of any single sample, implying that the association was reliable. The forest plot of meta-analysis in Asians is shown in Figure 1A. Although it did not achieve genome-wide level of statistical significance, the observed OR (0.928) of rs11191580 in Asian samples was in agreement with the European GWAS results on this SNP, suggesting that it was likely a real common risk SNP for schizophrenia in general populations.

We then conducted the association analysis of rs11191580 with brain structures in an Asian sample of 294 healthy subjects. The results showed that rs11191580 was significantly associated with the total white matter volume ($p = 0.019$), and individuals carrying the non-risk allele (C) of rs11191580 tended to have larger white matter volumes (Figure 1B). These data implied that rs11191580 likely conferred risk of schizophrenia susceptibility through affecting brain development, especially formation/development of white matter.

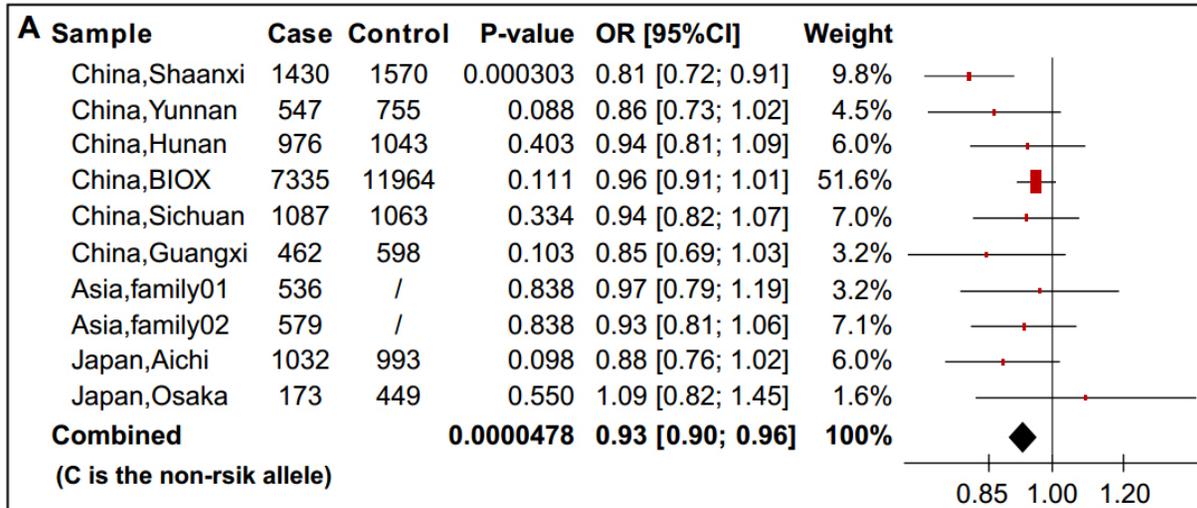


Figure 1 (A) Forest plot for meta-analysis of rs11191580 [C allele] with schizophrenia in Asian populations

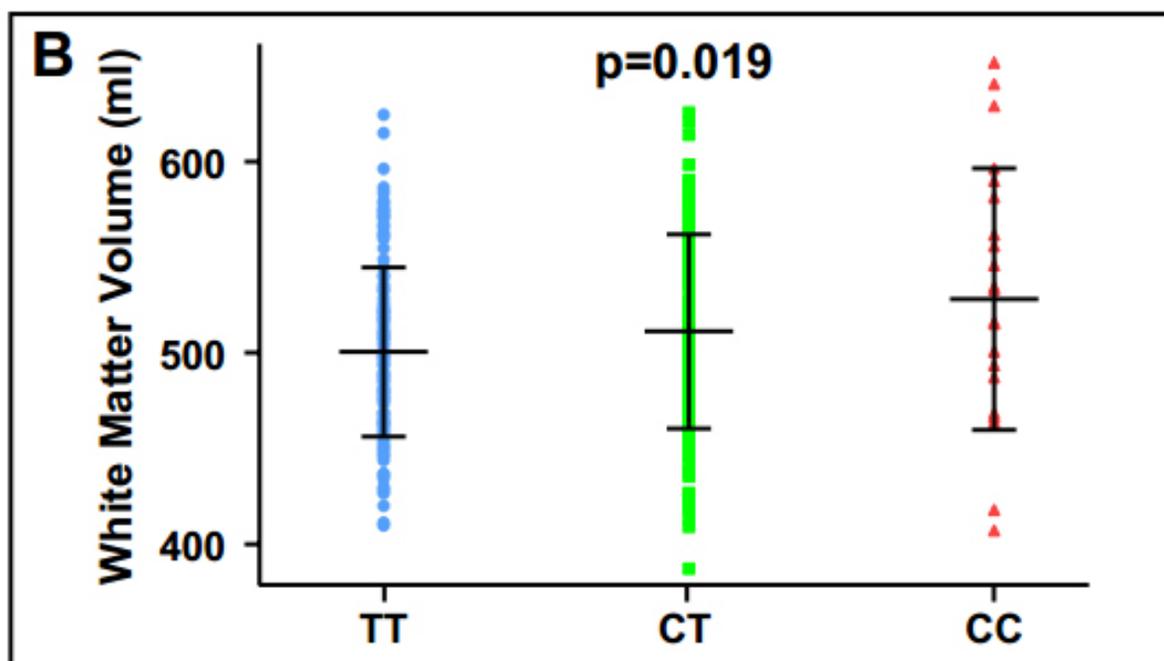
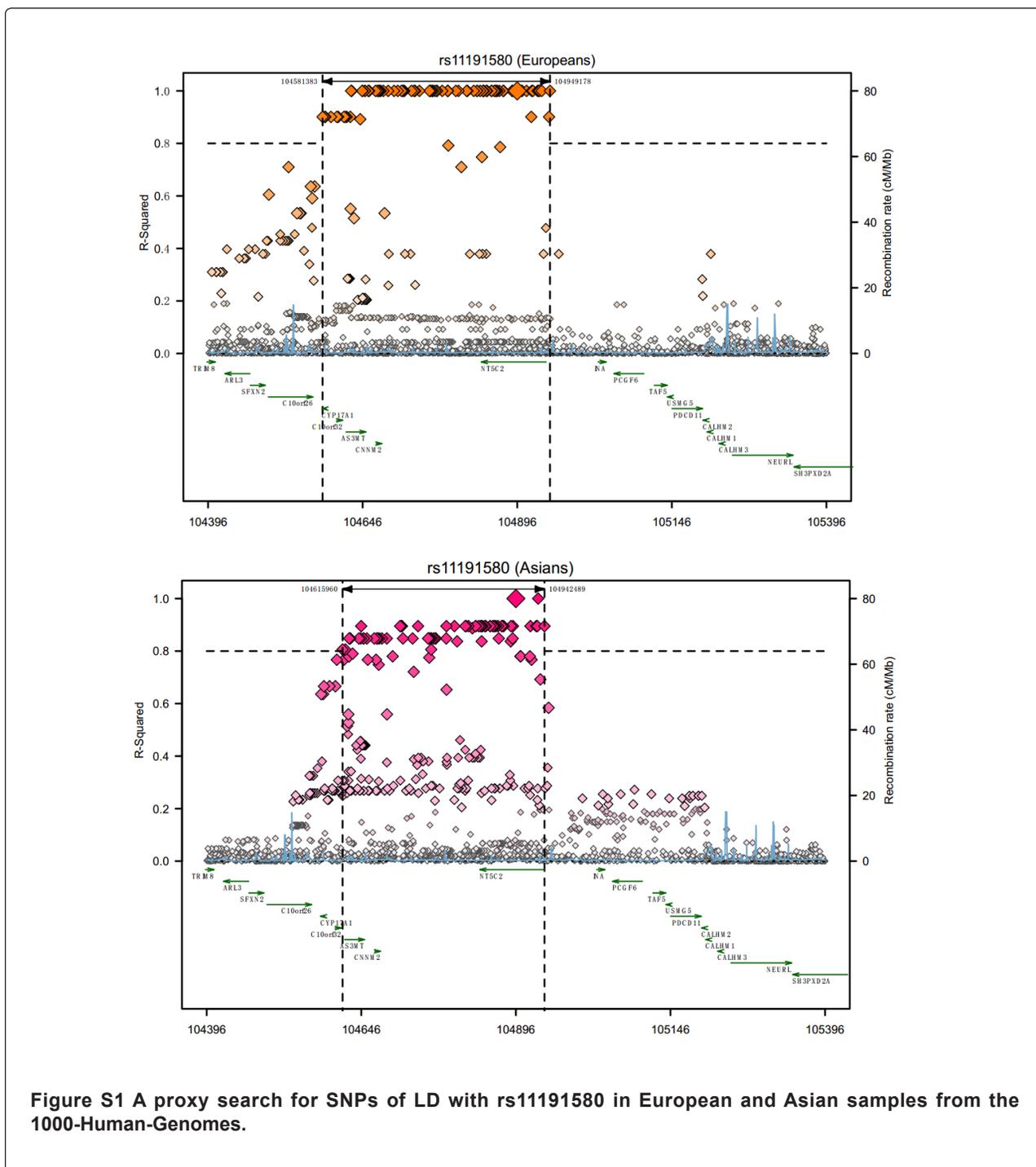


Figure 1 (B) Association of rs11191580 with white matter volume in Asian populations.

In genetic association studies, an associated SNP most likely points to a larger region of correlated variants exhibiting high degree of linkage disequilibrium (LD), making it difficult to precisely identify the causal variant. To investigate if there were SNPs linked with rs11191580, we explored the LD between rs11191580 and surrounding SNPs. A proxy search for SNPs of LD with rs11191580 was performed on the SNAP website with European

and Asian panels from the 1000-Human-Genomes (pilot 1) dataset. This search found that there were extensive SNPs in relative high LD ($r^2 > 0.80$) with rs11191580, spanning a large genomic region covering CYP17A1, BORCS7, AS3MT, CNNM2 and NT5C2 (Figure S1). Thus we could not exclude the possibility that rs11191580 was only a marker reflecting the association signal due to its LD with the causative variant.

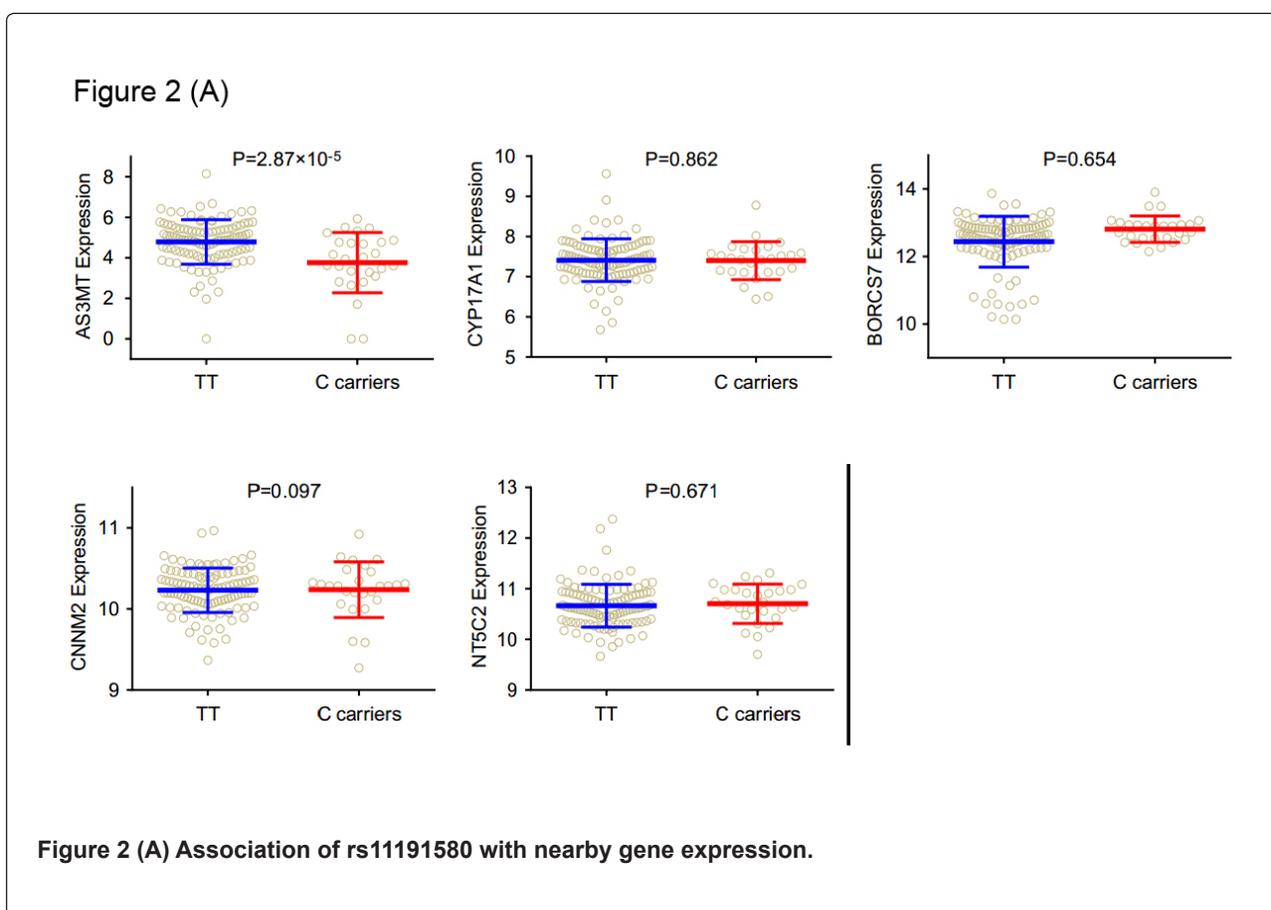


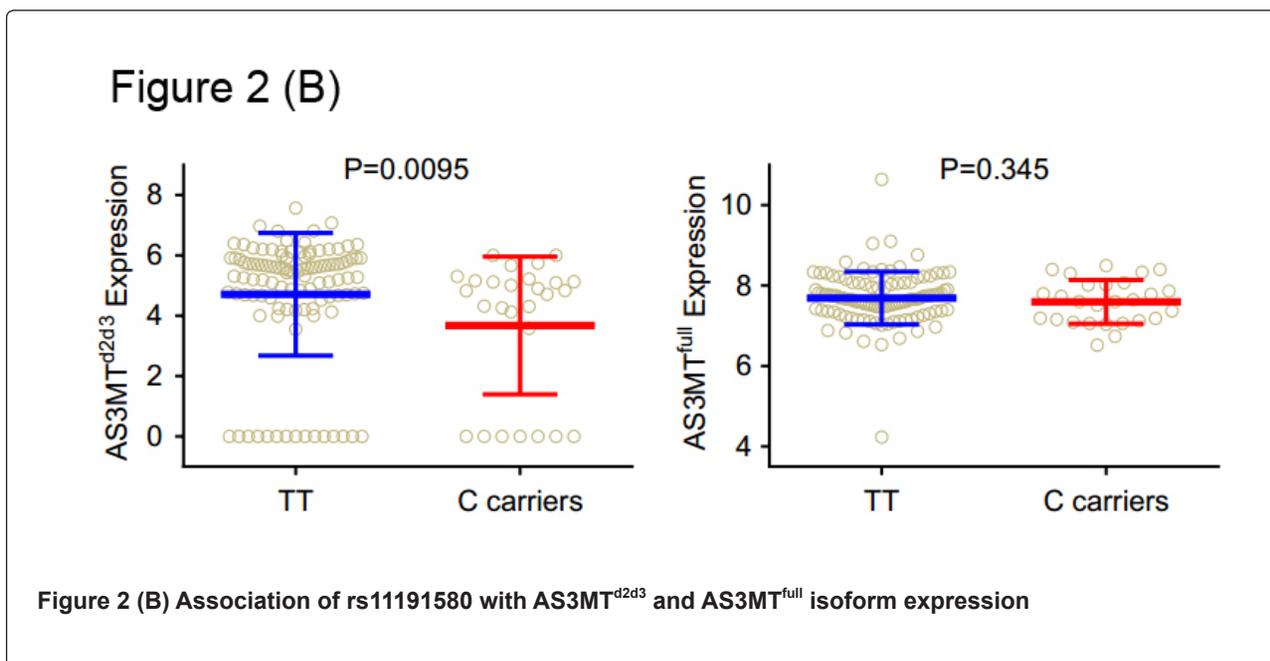
3.2 Rs11191580 was Associated with mRNA Expression of a Human-Specific Isoform of AS3MT

The associations of rs11191580 with schizophrenia lent statistical support to the involvement of this genomic region in the risk of the disorder. However, these findings did not identify the underlying molecular mechanism. To test the effects of the risk SNPs on the expression of nearby genes, we used a RNA-seq sample in the dorsolateral prefrontal cortex (DLPFC) of European healthy controls (n = 135). As shown in Figure S1, rs11191580 and its high LD SNPs spanned CYP17A1, BORCS7, AS3MT, CNM2 and NT5C2 in genome, therefore

we focused on these five genes.

The expression quantitative trait loci (eQTL) analysis on the gene level found that rs11191580 was only associated with the expression of AS3MT but not the other four genes in this genomic region ($p = 2.87 \times 10^{-5}$, Figure 2A). A recent study suggested that there was alternative splicing within the AS3MT gene to result in two primary isoforms, one was the full length AS3MT (named AS3MT^{full}), and the other was a truncated isoform missing exon 2 and 3 compared with AS3MT^{full} (named AS3MT^{d2d3}).^[26] The transcript level analysis found that rs11191580 was only associated with the human unique isoform AS3MT^{d2d3} ($p = 0.0095$, Figure 2B).^[26]





3.3 Pleiotropic Effects of rs1191580 on Complex Human Traits and Diseases in Europeans and Asians

Accumulating lines of evidence have suggested substantial overlap of risk factors between schizophrenia and non-psychiatric disorders such as blood pressure, cardiovascular diseases (CAD) and hypertension from clinical, epidemiological and genetic perspectives. We then investigated

if rs1191580 was also associated with these complex traits and diseases in both European and Asian populations. Interestingly, rs1191580 was significantly associated with blood pressure and CAD in both European and Asian populations (Table 1), and the schizophrenia non-risk [C] allele reduced risk of hypertension and CAD, which was in concordant with higher prevalence of hypertension and CAD in patients with schizophrenia across previous studies.

Table 1. Pleiotropic effects of rs1191580 on complex diseases and traits

Study	Ethnics	Traits	Sample size		Allele	Freq	Beta coefficient	P-value
			Initial	Replication				
PGC2 ^[3]	Mixed	Schizophrenia	>81,080	/	C	0.086	0.864 (OR)	7.07E-16
Deloukas et al. ^[28]	Mixed	CAD (Stage I)	83,595	/	C	0.110	0.930 (OR)	3.43E-05
	Mixed	CAD (Stage II)	107,262	/	C	0.107	0.950 (OR)	3.91E-03
Kato et al. ^[31]	Asian	DBP	19,457	21,858	C	0.260	-0.680 mm Hg	6.60E-12
	Asian	SBP	19,457	21,858	C	0.260	-1.180 mm Hg	3.90E-17
Ehret et al. ^[29]	European	DBP	67,577	/	C	0.086	-0.506 mm Hg	3.69E-06

	European	SBP	67,592	/	C	0.086	-1.058 mm Hg	9.16E-10
	European	Hypertension	69,395	133,361	C	0.086	-0.097 unit	1.40E-05
Wain et al. ^[30]	European	MAP	28,010	/	C	0.086	-0.602 mm Hg	3.29E-09
	European	PP	71,666	/	C	0.086	-0.579 mm Hg	2.19E-08

Note: CAD, coronary artery disease; DBP, Diastolic blood pressure; SBP, Systolic blood pressure; MAP, Mean Arterial Pressure (a weighted average of SBP and DBP); PP, Pulse pressure (the difference between SBP and DBP, a measure of stiffness of the main arteries); OR, odds ratio

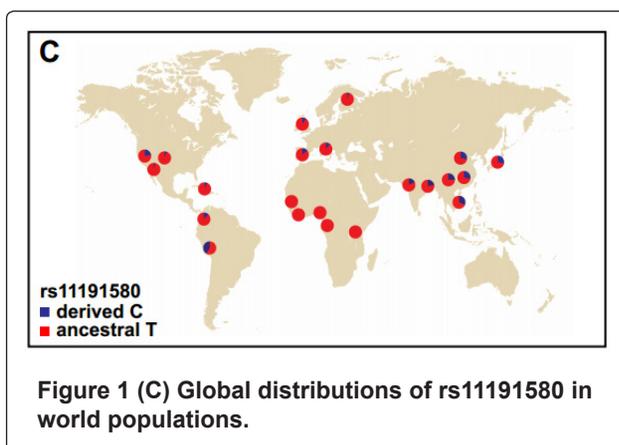
PGC2 is the abbreviation of Schizophrenia Working Group of the Psychiatric Genomics Consortium.

Together with the results of eQTL analysis, these data suggested that the human-unique AS3MT^{d2d3} isoform might be also involved in the development of metabolic syndromes such as CAD and hypertension, but further studies are needed.

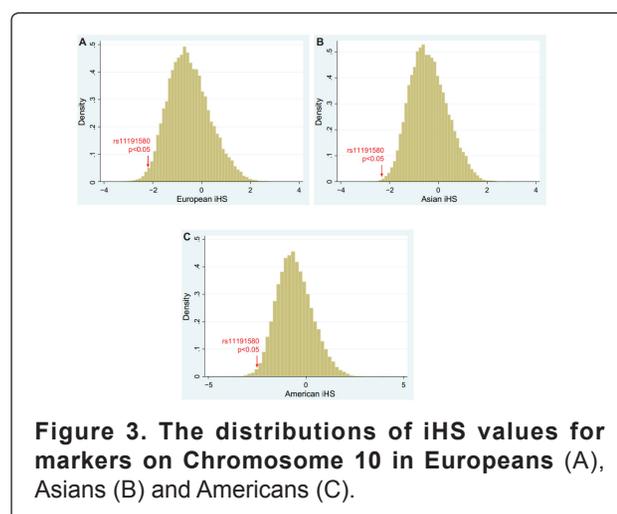
3.4 Evidence of Recent Positive Selection on rs11191580 in Europeans and Asians

Pleiotropic effects of a genomic locus on diverse phenotypes might be a reflection of natural selection (such example can be referred to rs13107325 in SLC39A8^[11]). To test this hypothesis, we performed population genetic analysis of rs11191580 using the genotype data from 1000-Human-Genome.

The geographic distribution analysis found that the frequencies of the derived allele (i.e., schizophrenia non-risk C-allele) of rs11191580 were maintained at ~ 25 % in Asians and ~ 10 % in Europeans, but totally absent in Africans. These data indicated a clear regional distribution pattern and a relatively young origin of this SNP (Figure 1C).



The regional enrichment of C-allele of rs11191580 in populations outside Africa mainland can be explained either by population substructure due to random genetic drift, or by recent positive selections in regional European and Asian populations. Long-range haplotype (LRH) tests can detect partial selective sweeps, particularly with allele frequencies as low as ~ 10 %. Therefore, we calculated the integrated haplotype score (iHS) (a statistic based on the extent of decay of LD surrounding a variant subjected to natural selection) of rs11191580 using the LRH tests. The iHS analysis of individuals from 1000-Human-Genome revealed that the derived C-allele experienced recent positive selection in Europeans (iHS = -2.16, $p < 0.05$), Asians (iHS = -2.22, $p < 0.05$) and Americans (iHS = -2.29, $p < 0.05$). In Africans, rs11191580 was monomorphic and had no iHS value. We also calculated the iHS of extensive common SNPs ($N > 600$) in chromosome 10, and found that rs11191580 was among those with the extremely 5 % lowest iHS values (Figure 3A-C). These data supported our hypothesis that rs11191580 underwent recent positive selection in European and Asian populations.



3.5 Potential Selection Pressures of Positive Selection

The association between rs11191580 with diverse human phenotypes and its recent positive selection in both Europeans and Asians are intriguing. The derivation of the new allele reduces risk of both schizophrenia and cardiovascular diseases in Europeans and Asians. Considering that modern humans migrated out of Africa mainland about 40,000 years ago,^[45] we speculated that this positive selection occurred after the arrival of modern human ancestors at Europe and Asia mainland given the evidence for a more recent positive selection event (< 30,000 years old) revealed by the LRH tests. Also, since the positive selection acted on both European and Asian populations but not Africans, the selection pressure might be related to historical events shared by Europe and Asia but not Africa.

3.5.1 Hypothesized selection pressure related to obesity

To our knowledge, one of the biggest environmental changes when modern humans migrated out of Africa mainland was the decreased ambient temperature. Considering that Europe and Asia continents were much colder than Africa mainland, the new alleles (e.g., rs11191580 C-allele) probably arose when modern humans migrated to the changed environment to help promote energy intake and expenditure to maintain thermal homeostasis (i.e., an optimal body temperature that is most often above their ambient temperature) in the colder climate. However, the same alleles also contributed to the contemporary increase in obesity rate. Notably, this hypothesis is supported by the significant associations between rs11191580 with obesity and BMI in both Europeans and Asians (Table 2), with the positive selected allele (the schizophrenia non-risk C-allele) increased risk of obesity. The positive selected (C) allele is also associated with higher caloric intake from protein (Table 2), which is consistent with its role in obesity and further supports our speculation of the potential pressure of the positive selection on this genomic locus.

Table 2. Effects of rs11191580 on obesity and energy intake

Study	Ethnics	Traits	Sample size		Allele	Freq	Beta coefficient	P-value
			Initial	Replication				
Wen et al. ^[34]	Asian	Obesity	84,040	14,843	C	0.270	1.050 (OR)	3.83E-04
	Asian	BMI	84,040	14,843	C	0.270	0.0295	3.83E-08
Speliotes et al. ^[33]	European	BMI	126,528		C	0.058	0.0345	9.19E-06
Present study	European	Energy intake#	71,893		C	0.088	0.1281	3.12E-02

Note: BMI, body mass index; OR, odds ratio

#Percentage of total caloric intake from fat; statistical analysis was performed by covaring sex, age and sub-population stratification.

3.5.2 Hypothesized selection pressure related to hypertension

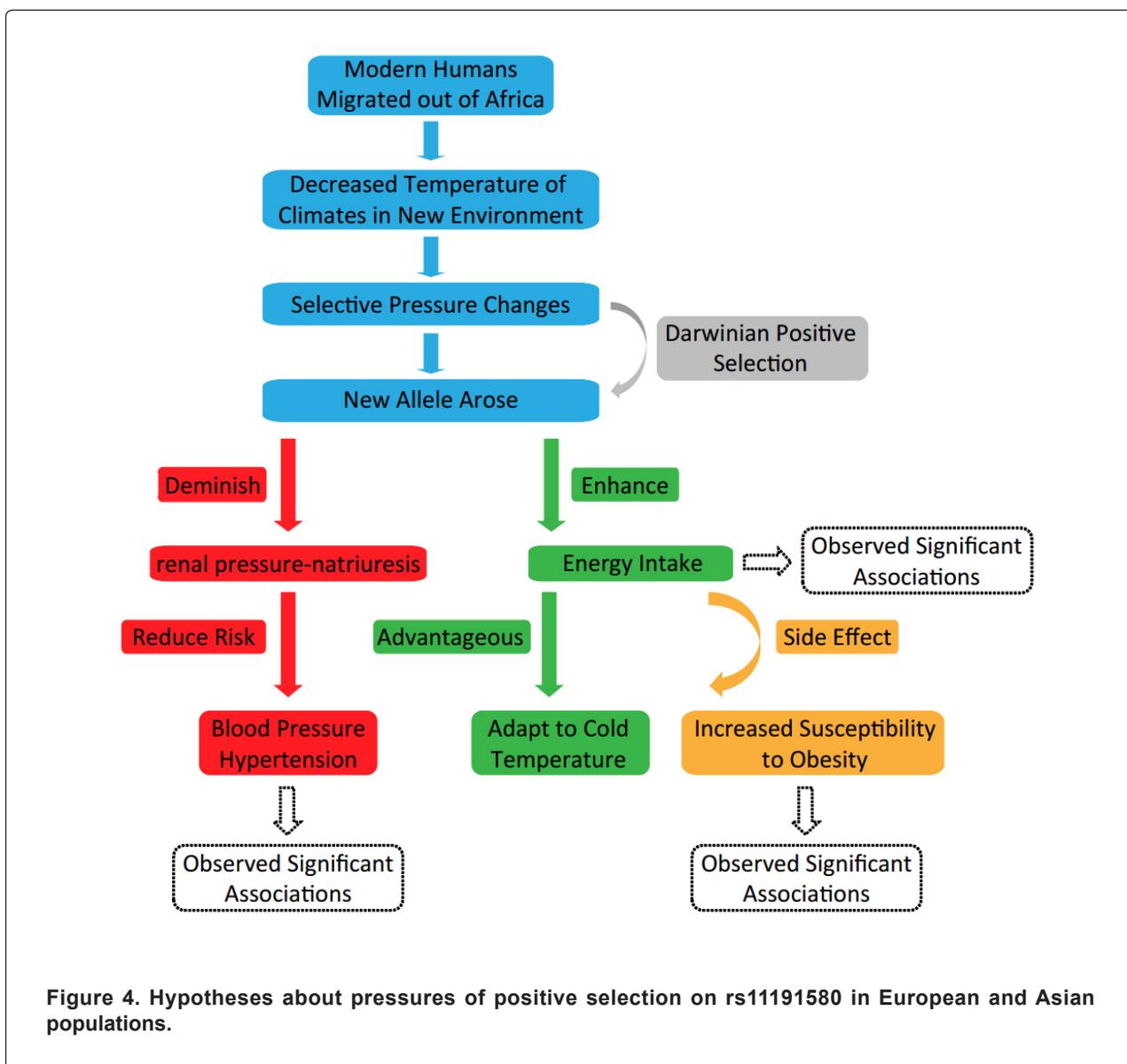
It is known that when entering a new environment, e.g. another continent, some ancestral alleles which had been beneficial in Africa may become deleterious.^[46] Under such conditions, some new alleles arose because of positive selection and

protected humans from maladaptation to the new environment and development of diseases.^[47] In this study, it is compelling to observe that the derived C-allele of rs11191580 reduces blood pressure and risk of hypertension in both Europeans and Asians. One of the most prevalent evolutionary hypotheses about hypertension (and blood pressure) is the

“sodium hypothesis”,^[46,48] which argues that natural selection for sodium conservation in the hot, dry savannah climate (e.g., Africa) into which humans first emerged could have resulted in sodium avidity that today is maladaptive. As humans migrated out of Africa, selection pressure for sodium conservation was expected to decline in the cooler and wetter climates of the northern latitudes (e.g., Europe and Asia). As a result, ancestral sodium-conserving genotypes (the called “thrifty genotype”) were expected to persist as the result of genetic drift, and might lead to an increased risk of hypertension in a changed environment of sodium abundance (e.g., Europe and Asia) where they changed the set point of the renal pressure-natriuresis to a higher blood pressure range. In this case, the new alleles arose and underwent positive selection in the cooler and

wetter climates (e.g., Europe and Asia), leading to reduced risk of hypertension. This hypothesis has been validated by previous observations on several hypertension candidate genes (e.g., CYP3A and AGT),^[49,50] and rs11191580 in 10q24.33 region is likely another example of such positive selection. Therefore, the protective effects of C-allele in rs11191580 from hypertension and blood pressures may reflect another pressure for this positive selection.

Taken together, we propose the intercorrelated relationship between positive selection of host alleles, hypertension, energy intake, and the human evolutionary histories (Figure 4). The positive selected risk C-allele might be beneficial for humans to better adapt to the Europe and Asia environment.



4 DISCUSSION

Chromosome 10q24.33 is one of the most significant loci for schizophrenia in recent GWAS among European populations,^[3] but the associations in other ethnic groups are relatively unclear. Using independent samples from multiple Asian area, we confirm that rs11191580 in 10q24.33 is also associated with schizophrenia in this distinct population. Although it did not achieve the genome-wide level of statistical significance, this observation was reasonable considering the limited sample size. We also show that the risk SNP is nominally associated with the brain structure variation, adding further evidence for the involvement of this genomic region in brain development and schizophrenia pathogenesis. We further confirm that rs11191580 is associated with the expression of a human-unique isoform of AS3MT,^[26] which is consistent with the previous study.

Interestingly, the schizophrenia risk SNP is also associated with hypertension and CAD in both European and Asian populations. After analyses of this SNP with multiple metabolic malfunction related conditions, we observe the same direction of allelic effects on all conditions, suggesting pleiotropic effects of this genomic sequence on diverse human phenotypes. This may explain partially the observed comorbidity of schizophrenia and hypertension/CAD. It is worth noting that the schizophrenia non-risk C-allele at rs11191580 reduces risk of several metabolic syndromes (e.g., hypertension and CAD), which is also consistent with the higher prevalence of metabolic syndromes among schizophrenic individuals.^[51] Moreover, the significant associations of rs11191580 with AS3MT^{d2d3} expression implies that AS3MT^{d2d3} may be a potential drug target for both metabolic syndromes and schizophrenia.

The pleiotropic effects on diverse human phenotypes lead us to perform the population genetic analysis on this genomic locus to understand its molecular evolution mechanisms. We have observed recent positive selection on rs11191580 that results in an increased frequency of the derived C-allele (schizophrenia non-risk allele) in Europeans and Asians. Considering that modern humans migrated out of Africa mainland about 40,000 years ago, we speculate that this positive selection occurred after the arrival of modern human ancestors at Europe and Asia continents. Since the long-range haplotype (LRH) tests find evidence for a more recent positive selection event (< 30,000 years old), the pressure of this selection may be related to common historical events in Europe and Asia but not Africa, such as

the colder climate. However, cautions should also be taken in interpreting these results, i.e., although we identified significant associations of rs11191580 with multiple phenotypes in distinct populations as well as the signature of recent positive selection on this locus, whether rs11191580 is the causative variant remains inconclusive, and the positive selection may act on a large genomic region due to the observed high LD in both populations. To nail down the exact risk variant, further studies are necessary.

Given the clear positive selection on this genomic locus, we propose two possible selection pressures. The first one is related to energy intake and its by-product obesity, while the other one is associated with blood pressure and hypertension. However, we do not identify selection pressure related to schizophrenia as the newly derived allele of rs11191580 confers reduced risk of schizophrenia in both populations. This phenomenon suggests that Darwinian positive selection may also protect humans from schizophrenia, but the relevant mechanism awaits further investigation. Also, it should be noted that the lifetime prevalence of schizophrenia does not differ significantly between continental populations. Since schizophrenia is a complex disorder involving thousands of alleles (may be different sets of risk alleles between populations) with limited effect of each, the combined contributions of all these alleles probably result in the similar prevalence of the disorders worldwide. We thus presume that positive selection on schizophrenia alleles would not be restricted within single populations, and there should be other risk alleles under positive selection in different populations (such as African), which are yet to be identified.

In summary, we have identified a SNP rs11191580 showing pleiotropic effects with diverse human phenotypes in Europeans and Asians, and we detected evidence of positive selection on the derived allele in both populations. Our evolutionary and genetic analyses may offer a unique and powerful opportunity to bring approaches from different perspectives together to discover how and why human diseases risks have evolved.^[52]

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This work was supported by CAS Pioneer Hundred Talents Program (to M.L.). The authors are deeply grateful for Miaoxin Li and Pak-Chung Sham (The University of Hongkong, China) for sharing their results of Sichuan samples. The authors also thank CHARGE Nutrition Working Group and DietGen Consortium for sharing their results about energy intake.

REFERENCES

- Saha S, Chant D, Welham J, McGrath J. A systematic review of the prevalence of schizophrenia. *PLoS Med* 2005; 2(5): e141.
- Sullivan PF, Kendler KS, Neale MC. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Archives of general psychiatry* 2003; 60(12): 1187-1192.
- Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 2014; 511(7510): 421-427.
- Carlner H, Collins PY, Cabassa LJ, McNallen A, Joestl SS, Lewis-Fernandez R. Prevalence of cardiovascular risk factors among racial and ethnic minorities with schizophrenia spectrum and bipolar disorders: a critical literature review. *Compr Psychiatry* 2014; 55(2): 233-247.
- Gladigau EL, Fazio TN, Hannam JP, Dawson LM, Jones SG. Increased cardiovascular risk in patients with severe mental illness. *Intern Med J* 2014; 44(1): 65-69.
- Wagner GP, Zhang J. The pleiotropic structure of the genotype-phenotype map: the evolvability of complex organisms. *Nat Rev Genet* 2011; 12(3): 204-213.
- Wang YQ, Su B. Molecular evolution of microcephalin, a gene determining human brain size. *Human molecular genetics* 2004; 13(11): 1131-1137.
- Wang JK, Li Y, Su B. A common SNP of MCPH1 is associated with cranial volume variation in Chinese population. *Human molecular genetics* 2008; 17(9): 1329-1335.
- Luo XJ, Mattheisen M, Li M, Huang L, Rietschel M, Borglum AD *et al*. Systematic Integration of Brain eQTL and GWAS Identifies ZNF323 as a Novel Schizophrenia Risk Gene and Suggests Recent Positive Selection Based on Compensatory Advantage on Pulmonary Function. *Schizophr Bull* 2015; 41(6): 1294-1308.
- Keller MC, Miller G. Resolving the paradox of common, harmful, heritable mental disorders: which evolutionary genetic models work best? *Behav Brain Sci* 2006; 29(4): 385-404; discussion 405-352.
- Li M, Wu DD, Yao YG, Huo YX, Liu JW, Su B *et al*. Recent Positive Selection Drives the Expansion of a Schizophrenia Risk Nonsynonymous Variant at SLC39A8 in Europeans. *Schizophrenia bulletin* 2016; 42(1): 178-190.
- Guan F, Wei S, Feng J, Zhang C, Xing B, Zhang H *et al*. Association study of a new schizophrenia susceptibility locus of 10q24.32-33 in a Han Chinese population. *Schizophrenia research* 2012; 138(1): 63-68.
- Li M, Wang Y, Zheng XB, Ikeda M, Iwata N, Luo XJ *et al*. Meta-analysis and brain imaging data support the involvement of VRK2 (rs2312147) in schizophrenia susceptibility. *Schizophrenia research* 2012; 142(1-3): 200-205.
- Ma L, Wu DD, Ma SL, Tan L, Chen X, Tang NL *et al*. Molecular evolution in the CREB1 signal pathway and a rare haplotype in CREB1 with genetic predisposition to schizophrenia. *Journal of psychiatric research* 2014; 57: 84-89.
- Li Z, Xiang Y, Chen J, Li Q, Shen J, Liu Y *et al*. Loci with genome-wide associations with schizophrenia in the Han Chinese population. *Br J Psychiatry* 2015; 207(6): 490-494.
- Wong EH, So HC, Li M, Wang Q, Butler AW, Paul B *et al*. Common variants on Xq28 conferring risk of schizophrenia in Han Chinese. *Schizophrenia bulletin* 2014; 40(4): 777-786.
- Li Z, Jiang J, Long J, Ling W, Huang G, Guo X *et al*. The rs11191580 variant of the NT5C2 gene is associated with schizophrenia and symptom severity in a South Chinese Han population: evidence from GWAS. *Rev Bras Psiquiatr* 2016; 0.
- Takahashi S, Glatt SJ, Uchiyama M, Faraone SV, Tsuang MT. Meta-analysis of data from the Psychiatric Genomics Consortium and additional samples supports association of CACNA1C with risk for schizophrenia. *Schizophrenia research* 2015; 168(1-2): 429-433.
- Aberg KA, Liu Y, Bukszar J, McClay JL,

- Khachane AN, Andreassen OA *et al*. A comprehensive family-based replication study of schizophrenia genes. *JAMA Psychiatry* 2013; 70(6): 573-581.
20. Saito T, Kondo K, Iwayama Y, Shimasaki A, Aleksic B, Yamada K *et al*. Replication and cross-phenotype study based upon schizophrenia GWASs data in the Japanese population: Support for association of MHC region with psychosis. *American journal of medical genetics Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics* 2014; 165(5): 421-427.
 21. Ohi K, Hashimoto R, Yamamori H, Yasuda Y, Fujimoto M, Umeda-Yano S *et al*. The impact of the genome-wide supported variant in the cyclin M2 gene on gray matter morphology in schizophrenia. *Behav Brain Funct* 2013; 9: 40.
 22. Kazeem GR, Farrall M. Integrating case-control and TDT studies. *Ann Hum Genet* 2005; 69(Pt 3): 329-335.
 23. Li M, Ge T, Feng J, Su B. SLC6A15 rs1545843 and Depression: Implications From Brain Imaging Data. *Am J Psychiatry* 2013; 170(7): 805.
 24. Patenaude B, Smith SM, Kennedy DN, Jenkinson M. A Bayesian model of shape and appearance for subcortical brain segmentation. *Neuroimage* 2011; 56(3): 907-922.
 25. Zhang Y, Brady M, Smith S. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE transactions on medical imaging* 2001; 20(1): 45-57.
 26. Li M, Jaffe AE, Straub RE, Tao R, Shin JH, Wang Y *et al*. A human-specific AS3MT isoform and BORCS7 are molecular risk factors in the 10q24.32 schizophrenia-associated locus. *Nat Med* 2016; 22(6): 649-656.
 27. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 2009; 25(9): 1105-1111.
 28. Deloukas P, Kanoni S, Willenborg C, Farrall M, Assimes TL, Thompson JR *et al*. Large-scale association analysis identifies new risk loci for coronary artery disease. *Nature genetics* 2013; 45(1): 25-33.
 29. Ehret GB, Munroe PB, Rice KM, Bochud M, Johnson AD, Chasman DI *et al*. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* 2011; 478(7367): 103-109.
 30. Wain LV, Verwoert GC, O'Reilly PF, Shi G, Johnson T, Johnson AD *et al*. Genome-wide association study identifies six new loci influencing pulse pressure and mean arterial pressure. *Nature genetics* 2011; 43(10): 1005-1011.
 31. Kato N, Takeuchi F, Tabara Y, Kelly TN, Go MJ, Sim X *et al*. Meta-analysis of genome-wide association studies identifies common variants associated with blood pressure variation in east Asians. *Nature genetics* 2011; 43(6): 531-538.
 32. Stunkard AJ, Foch TT, Hrubec Z. A twin study of human obesity. *Jama* 1986; 256(1): 51-54.
 33. Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU *et al*. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nature genetics* 2010; 42(11): 937-948.
 34. Wen W, Zheng W, Okada Y, Takeuchi F, Tabara Y, Hwang JY *et al*. Meta-analysis of genome-wide association studies in East Asian-ancestry populations identifies four new loci for body mass index. *Human molecular genetics* 2014.
 35. Willett W. *Nutritional epidemiology*, vol. 40. Oxford University Press 2013.
 36. Tanaka T, Ngwa JS, van Rooij FJ, Zillikens MC, Wojczynski MK, Frazier-Wood AC *et al*. Genome-wide meta-analysis of observational studies shows common genetic variants associated with macronutrient intake. *The American journal of clinical nutrition* 2013; 97(6): 1395-1402.
 37. Chu AY, Workalemahu T, Paynter NP, Rose LM, Giulianini F, Tanaka T *et al*. Novel locus including FGF21 is associated with dietary macronutrient intake. *Human molecular genetics* 2013; 22(9): 1895-1902.
 38. Genomes Project Consortium, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM *et al*. A map of human genome variation from population-scale sequencing. *Nature* 2010; 467(7319): 1061-1073.
 39. Szpiech ZA, Hernandez RD. selscan: An Efficient Multithreaded Program to Perform EHH-Based Scans for Positive Selection. *Molecular biology and evolution* 2014; 31: 2824-2827.
 40. Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsepas C *et al*. Genome-wide

- detection and characterization of positive selection in human populations. *Nature* 2007; 449(7164): 913-918.
41. Voight BF, Kudravalli S, Wen X, Pritchard JK. A map of recent positive selection in the human genome. *PLoS biology* 2006; 4(3): e72.
 42. Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ, Schaffner SF *et al*. Detecting recent positive selection in the human genome from haplotype structure. *Nature* 2002; 419(6909): 832-837.
 43. Schizophrenia Psychiatric Genome-Wide Association Study Consortium. Genome-wide association study identifies five new schizophrenia loci. *Nature genetics* 2011; 43(10): 969-976.
 44. Ripke S, O'Dushlaine C, Chambert K, Moran JL, Kahler AK, Akterin S *et al*. Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nature genetics* 2013; 45(10): 1150-1159.
 45. Underhill PA, Kivisild T. Use of y chromosome and mitochondrial DNA population structure in tracing human migrations. *Annu Rev Genet* 2007; 41: 539-564.
 46. Weder AB. Evolution and hypertension. *Hypertension* 2007; 49(2): 260-265.
 47. Hancock AM, Witonsky DB, Gordon AS, Eshel G, Pritchard JK, Coop G *et al*. Adaptations to climate in candidate genes for common metabolic disorders. *PLoS Genet* 2008; 4(2): e32.
 48. Gleibermann L. Blood pressure and dietary salt in human populations. *Ecology of food and nutrition* 1973; 2(2): 143-156.
 49. Nakajima T, Wooding S, Sakagami T, Emi M, Tokunaga K, Tamiya G *et al*. Natural selection and population history in the human angiotensinogen gene (AGT): 736 complete AGT sequences in chromosomes from around the world. *American journal of human genetics* 2004; 74(5): 898-916.
 50. Thompson EE, Kuttub-Boulos H, Witonsky D, Yang L, Roe BA, Di Rienzo A. CYP3A variation and the evolution of salt-sensitivity variants. *American journal of human genetics* 2004; 75(6): 1059-1069.
 51. Papanastasiou E. The prevalence and mechanisms of metabolic syndrome in schizophrenia: a review. *Ther Adv Psychopharmacol* 2013; 3(1): 33-51.
 52. Crespi BJ. The emergence of human-evolutionary medical genomics. *Evol Appl* 2011; 4(2): 292-314.