Confirmation of Genomewide Association Signals in Chinese Han Population Reveals Risk Loci for Ischemic Stroke

Hu Ding, MD; Yujun Xu, MD; Xunna Bao, MD; Xiaojing Wang, PhD; Guanglin Cui, MD; Wei Wang, MD, PhD; Rutai Hui, MD, PhD; Dao Wen Wang, MD, PhD

Background and Purpose—The first genomewide association study of ischemic stroke in whites has identified multiple susceptibility loci. We confirmed this study by examining associations with ischemic stroke in a Chinese Han population.

Methods—Twenty-five common variants were genotyped in a relatively large sample size including 1123 subjects with ischemic stroke cases (thrombosis stroke=716, lacunar infarction=407) and 557 normal control subjects. The association analyses were performed at both single nucleotide polymorphism and haplotype levels. False discovery rate q value method was applied for multiple testing corrections.

Results—rs11052413, a intergenic single nucleotide polymorphism, was most significantly associated with ischemic stroke independent of traditional cardiovascular risk factors in additive (OR=1.51, 95% CI=1.19 to 1.92, \( P=7.4\times10^{-4} \), \( q=0.018 \)) and dominant models (OR=1.59, 95% CI=1.20 to 2.08, \( P=9.2\times10^{-4} \), \( q=0.023 \)). In addition, both ZNF650 rs10204475 and intergenic single nucleotide polymorphism rs10486776 were associated with ischemic stroke as well as independent of traditional cardiovascular risk factors in dominant models (OR=1.47, 95% CI=1.12 to 1.96, \( P=0.005 \), \( q=0.040 \) and OR=1.53, 95% CI=1.15 to 2.02, \( P=0.003 \), \( q=0.036 \), respectively). No significant results were found in stroke subtype analysis after multiple corrections.

Conclusion—Our study confirmed previously reported associations between ischemic stroke and rs11052413, rs10486776, and ZNF 650 rs10204475 in a Chinese Han population. The mechanism whereby the genetic variants exert their effects on ischemic stroke remains to be further elucidated. (Stroke. 2010;41:177-180.)

Key Words: genetics ■ ischemic stroke ■ polymorphisms

Stroke is a complex trait that is assumed to be caused by both genetic and environmental factors as well as their interactions.\(^1\) Genetic factors have been defined as an important risk contributor to the pathogenesis of ischemic stroke,\(^2\) but the responsible molecular and genetic determinants remain largely unidentified.\(^3\) Over the past 2 decades, classical linkage approaches as well as genetic association studies have been disappointing, yielding inconsistent and nonreplicable findings.\(^4\) Since 2007, the advent of large-scale genomewide association studies has provided a dramatic and surprisingly successful new tool in finding consistent and replicable genetic markers of several complex diseases of adulthood. In the population of European origin, Matarin et al recently reported the first phase of a genomewide association analysis in patients with stroke.\(^5\) Consistent examination of an effect for multiple risk loci in various populations would provide stronger evidence of causality.\(^6\) Therefore, we carried out a case–control association study in our Chinese Han cohort to determine whether those significant association signals reported in white populations could be confirmed in Asians.

Materials and Methods

Study Population and Data Collection

This was a multicenter study for the assessment of risk factors for stroke and sponsored by the Ministry of Science and Technology of China. The study protocol was approved by the review board of the Ministry of Public Health, Ministry of Science and Technology of China and the ethics committees at all participating hospitals. Informed consent was obtained from all participants.

A total of 1388 subjects with stroke were recruited between November 2004 and January 2009 from 5 hospitals in Wuhan, China. Only 3 subtypes of stroke—cerebral atherosclerosis (atherothrombosis, \( n=716 \)), lacunar infarction (lacunar, \( n=407 \)), and intracerebral hemorrhage (\( n=275 \))—were included. Thus, 1123 patients with ischemic stroke were involved in this case–control association study. Confirmation of stroke was based on the results of strict neurological
examination, CT, or MRI according to the International Classification of Diseases (9th Revision, codes 430 to 438). Stroke was defined as a sudden onset of nonconvulsive and focal neurological deficit persisting for >24 hours. Cerebral thrombosis was diagnosed if the patient had the characteristics of cerebral cortical dysfunction and CT scan or MRI performed within 28 days of onset either showed an area of low attenuation in a region compatible with the clinical symptoms or failed to identify an alternative cause of the symptoms. Lacunar infarction was diagnosed if the patient had one of the characteristic clinical lacunar syndromes and CT scan or MRI showed either normal or a deep focal infarction in the brain stem or subcortical region with a diameter of <15 mm. Other types of stroke (transient ischemic attack, subarachnoid hemorrhage, embolic brain infarction, brain tumors, and cerebrovascular malformation) and severe systemic diseases such as pulmonary fibrosis, endocrine and metabolic disease (except diabetes mellitus), severe inflammatory diseases, autoimmune disease, tumors, and serious chronic diseases (eg, hepatic cirrhosis, renal failure) were excluded from this study. Subjects with cardiomyopathic stroke and documented atrial fibrillation were also excluded from our study.

Five hundred fifty-seven ethnically and geographically matched control subjects were randomly recruited either from healthy residents in the community (n=499 [89.6%]) or inpatients (n=58 [10.4%]) with minor illnesses by a set of recruitment questionnaires. All control subjects received a physical examination of neurological systems and were free of neurological and cardiovascular diseases following the same exclusion criteria as cases. Initially, 600 control subjects were recruited. Before data analysis, 43 control subjects were excluded for indefinite diagnosis or insufficient DNA. No significant differences were found in clinical characteristics between the included and excluded subjects.

**Single Nucleotide Polymorphism Selection**

In the previous genomewide association study by Matarin et al, a total of 27 variants were reported with susceptibility to ischemic stroke. However, 2 single nucleotide polymorphisms (SNPs) are very rare (rs13126803, minor allele frequency=0.023) or nonexistent (rs246341) in Asians according to the HapMap Project Phase II database. Thus, we selected the remaining 25 SNPs for genotyping in our study.

**DNA Isolation and Genotyping**

Genomic DNA was isolated from whole blood collected in K$_2$-EDTA tubes using the QG-Mini80 workflow with a DB-S kit (FUJIFilm Corporation, Tokyo, Japan) as instructed. DNA was quantified and diluted to a final concentration of 10 ng/μL.

All samples were genotyped using the Taqman 7900HT Sequence Detection System according to the manufacturer’s instructions. Each assay was carried out using 10 ng DNA in a 5-μL reaction consisting of a TaqMan universal polymerase chain reaction master mix (Applied Biosystems, Foster City, Calif), forward and reverse primers, and 6-carboxyfluorescein (FAM)- and 4,7,2-trichloro-7-phenyl-6-carboxyfluorescein (VIC)-labeled probes designed by Applied Biosystems (ABI Assays-on-Demand). Allelic discrimination was measured automatically using the Sequence Detection System 2.3 software (including confidence level 95%). A total of 10% of all genotypes were repeated in independent polymerase chain reactions to check for consistency and to ensure intraplate and interplate genotype quality control. No genotyping discrepancies were detected between the repeated samples. In addition, all the DNA samples for cases and control subjects were run in the same batch.

**Statistical Analysis**

The presence of Hardy-Weinberg equilibrium per SNP was tested by χ2 goodness-of-fit test. Multiple unconditional logistic regression was used to estimate OR and 95% CI after adjusting for covariates (gender, age, body mass index, hypertension, diabetes, hyperlipidemia, and smoking status) under both additive and dominant models. The false discovery rate (FDR) adjustment for multiple testing probability value is given by the q value method using QVALUE software (setting λ=0, FDR level=0.05).8 Haplotype frequencies for various SNP combinations were first estimated by HAPLOSTATS© (Version 1.2.1) in R statistical package and then verified using PHASE© (Version 2.1). The HAPLOSTATS program could help compute global score and haplotype-specific score probability values while allowing for adjusting covariates under the additive model using default settings. Power calculations were performed using QUANTO software program (Version 1.2.3).11 All other statistical analyses were performed with SPSS 13.0 (SPSS Inc, Chicago, Ill) in Windows (Microsoft Corp, Redmond, Wash) and SNPassoc function within the R statistical package.12

**Results**

The demographic details for the patients and control subjects are shown in Table 1. To demonstrate the genetic susceptibility of SNPs with ischemic stroke in the Chinese Han population, 25 SNPs were genotyped in 1123 subjects with ischemic stroke (thrombosis stroke=716, lacunar infarction=407) and 557 normal control subjects. All the SNPs conformed to Hardy-Weinberg equilibrium except rs32720, which significantly deviated from Hardy-Weinberg equilibrium in both control subjects and cases. Thus, 24 SNPs were selected for further analysis. The relevant information for all these SNPs is available from Table 2. Data were then analyzed using multiple unconditional logistic regressions with adjustment for covariates under additive and dominant models. For multiple testing corrections, we present the FDR q value, a measure of FDR expected for a given probability value in the follow-up analysis. All the corresponding FDR q values are available in Supplemental Table I (available at http://stroke.ahajournals.org). Our results revealed that an intergenic SNP, rs11052413, was most significantly associated with ischemic stroke independent of traditional cardiovascular risk factors in both additive (OR=1.51, 95% CI=1.19 to 1.92, P=7.4×10⁻⁴, q=0.018) and dominant models (OR=1.59, 95% CI=1.20 to 2.08, P=9.2×10⁻⁴, q=0.023). In addition, both ZNF650 rs10204475 and inter-

**Table 1. Clinical Characteristics of Cases and Control Subjects**

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Control Subjects</th>
<th>Ischemic Stroke</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>557</td>
<td>1123</td>
</tr>
<tr>
<td>Age, years</td>
<td>62.2±9.3</td>
<td>62.2±10.1</td>
</tr>
<tr>
<td>Men, %</td>
<td>62.1</td>
<td>66.1</td>
</tr>
<tr>
<td>Hypertension, %</td>
<td>19.2</td>
<td>69.2*</td>
</tr>
<tr>
<td>Diabetes, %</td>
<td>3.2</td>
<td>18.0*</td>
</tr>
<tr>
<td>Hyperlipidemia, %</td>
<td>20.9</td>
<td>28.7*</td>
</tr>
<tr>
<td>Smokers, %</td>
<td>37.3</td>
<td>52.6*</td>
</tr>
<tr>
<td>Body mass index, kg/m²</td>
<td>23.7±3.2</td>
<td>24.3±3.3*</td>
</tr>
<tr>
<td>High-density lipoprotein cholesterol, mmol/L</td>
<td>1.33±0.36</td>
<td>1.10±0.45*</td>
</tr>
<tr>
<td>Total cholesterol, mmol/L</td>
<td>4.56±1.70</td>
<td>4.53±1.32</td>
</tr>
</tbody>
</table>

*P values (P<0.01) for test of differences between cases and control subjects.
Discrimination between ischemic stroke independent of traditional cardiovascular risk factors is dominant models (OR=1.47, 95% CI=1.12 to 1.96, P=0.005, q=0.400 and OR=1.53, 95% CI=1.15 to 2.02, P=0.003, q=0.036, respectively).

To test the possible effect of ischemic stroke subtypes in detecting an association, we then reassessed the association between the same SNPs and 2 subtypes: thrombosis stroke and the lacunar infarction (Supplemental Tables II and III). It is of interest to note that the rs11052413 significantly associated with thrombosis stroke with the strongest risk; ZNF650 rs10204475 and rs10486776 were significantly associated with lacunar infarction with the strongest risk (nominal P<0.05). However, none of these results pass the significance threshold after multiple correction (FDR q value ranges 0.072 to 0.096). These negative results could be due to reduced sample size and statistical power for subtype analysis. We also performed haplotype analysis for associations between ischemic stroke and multiple SNPs. Unfortunately, none of the constructed haplotypes within KCNK17 or ASTN2 genes confers the risk for stroke (Supplemental Table IV).

After implementation of this study, we performed a post hoc statistical power analysis to verify whether the recruited samples can provide adequate power for identifying the association between SNPs with ischemic stroke based on our sample size. Assuming disease prevalence between 0.5% to 1%, our sample size can reach >90% power to detect a susceptibility locus with a genotypic relative risk >1.47 at the nominal Type I error rate <0.05 for SNPs with minor allele frequency >0.12 under both additive and dominant models. This power analysis indicates that our cohort sample size is sufficient to generate robust estimates in association analysis.

Discussion

The present study was carried out in a Chinese Han population to verify positive association signals identified in the first genomewide association study of ischemic stroke in whites. Our results revealed that ZNF650 rs10204475 and 2 intergenic SNPs, rs11052413 and rs10486776, were shown to be associated with ischemic stroke independent of traditional cardiovascular risk factors. Inconsistent with a previous report, all other positive association signals were not reproducible in our Chinese Han cohort. Notably, the 3 most significant SNPs reported by Matarin et al are within candidate gene KCNK17, which is worth further study. Data from Matarin et al and our group suggest that the frequencies of
risk alleles for 3 SNPs within KCNK17 (rs2395721, rs10947803, and rs10807204) in the white control population differ from the Chinese Han control population, 0.18 versus 0.37 for all the alleles. Therefore, studies conducted in the Chinese population should have a higher power to detect an association with stroke. However, neither haplotype analysis nor single maker adjustment analysis reproduced these associations. The absence of an association in the Chinese Han population may be explained by several possibilities: (1) lack of enough power to detect associations with alleles with a small effect (relative risk <1.12 in our paper); (2) heterogeneity among different ethnic groups; and (3) false-positive results in the previous genomewide association study. In fact, most of the genetic markers identified in the study published by Matarin et al did not achieve the threshold for genomewide significance (recommended probability values <10⁻⁸).

One intergenic SNP, rs11052413, was the most significant genetic marker associated with ischemic stroke in our study. Inspection of the UCSC Genome Browser (http://genome.ucsc.edu) and BLAST searches against the National Center for Biotechnology Information (www.ncbi.nih.gov/blast) revealed no annotated gene, spliced expressed sequence tags, or microRNAs containing this SNP. Therefore, the functional relevance of this intergenic SNP to ischemic stroke remains to be further elucidated. The identified SNP, rs10204475, in an intron of ZNF650 is probably not the causal variant; it is more likely that it is in linkage disequilibrium with the causal variants. However, according to the HapMap database, there is no known or obvious functional allele in linkage disequilibrium with this allele that explains these associations. Identification of the causal allele will ultimately require the resequencing and genotyping of samples from a large number of patients with ischemic stroke together with functional studies stratified by genotypes. Nevertheless, this study confirms the association of these SNPs previously identified in a European descent population and provides the first evidence of a cross-race susceptibility to ischemic stroke. Moreover, our present finding requires validation/confirmation in future prospective studies.

Acknowledgments
We thank Dr Tao Zhang, Dr Fei Ma, Dr Qian He, and Dr Lan Zhang for their efforts in recruiting patients and control subjects for this study. We want to especially acknowledge all the participants in this study.

Sources of Funding
This work was supported by grants from National “863” project (No. 2006AA02A406) and “973” projects (No. 2007CB512004 and No. 2006CB503801).

Disclosures
None.

References
Confirmation of Genomewide Association Signals in Chinese Han Population Reveals Risk Loci for Ischemic Stroke
Hu Ding, Yujun Xu, Xunna Bao, Xiaojing Wang, Guanglin Cui, Wei Wang, Rutai Hui and Dao Wen Wang

Stroke. 2010;41:177-180; originally published online November 12, 2009;
doi: 10.1161/STROKEAHA.109.567099
Stroke is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2009 American Heart Association, Inc. All rights reserved.
Print ISSN: 0039-2499. Online ISSN: 1524-4628

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://stroke.ahajournals.org/content/41/1/177

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Stroke can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Stroke is online at:
http://stroke.ahajournals.org//subscriptions/