

Fine-Mapping of ABO Gene Identifies Two Novel SNPs Associated with Large Artery Atherosclerotic Stroke in a Chinese Han Population

Hao Zhang^{1,2} · Zhizhong Zhang¹ · Jun Zhang¹ · Lili Xu¹ · Zusen Ye³ · Yonggang Hao³ · Biyang Cai¹ · Shuyu Zhou¹ · Keting Liu³ · Lingli Sun¹ · Wen Sun¹ · Yumeng Zhang⁴ · Huan Cai³ · Guihua Ni^{3,5} · Xinfeng Liu¹ · Gelin Xu¹

Received: 15 December 2015 / Accepted: 11 February 2016 / Published online: 29 February 2016
© Springer Science+Business Media New York 2016

Abstract Recently, a single nucleotide polymorphism (SNP) rs505922 in ABO gene was related to large artery atherosclerotic (LAA) stroke in Caucasian populations by genome-wide association study (GWAS). This study aimed to determine whether ABO gene polymorphisms are associated with LAA stroke in Chinese Han population. A case–control study was designed, and 644 patients with LAA stroke and 642 healthy controls were enrolled. Ten tagging SNPs (tSNPs) located in ABO gene were genotyped. Associations between genotypes and LAA stroke were analyzed with logistic regression model after adjustment of potential confounders. Although rs505922 was not associated with LAA stroke (TT genotype, adjusted OR = 1.32; 95 % CI, 0.94 to 1.87), two novel SNPs, rs8176668 (AT genotype, adjusted OR = 0.71; 95 % CI, 0.55 to 0.92) and rs2073824 (AA genotype, adjusted OR = 0.72; 95 % CI, 0.57 to 0.92), were associated with LAA stroke. Haplotype analysis indicated that haplotype TC (adjusted

OR = 0.72; 95 % CI, 0.54 to 0.95; $P = 0.018$) in block 1 and haplotype ACA in block 2 (OR = 0.73; 95 % CI, 0.58 to 0.91; $P = 0.005$) were associated with LAA stroke. Multifactor dimensionality reduction (MDR) analysis in the single-locus model indicated that rs2073824 was the most important attributor for predicting risk of LAA stroke. No significant SNP–SNP interactions among the tested SNPs were detected. The results indicated that the genetic variants in ABO gene may influence the risk of LAA stroke in Chinese Han population.

Keywords ABO · Atherosclerosis · Single nucleotide polymorphism · Stroke

Introduction

Stroke risks vary among ethnic populations, with much higher incidence and mortality reported in East Asian [1]. In China alone, 1.6 million people die of stroke every year, which rank stroke as the leading cause of death and adult disability in the country, and there are about 2.5 million new stroke patients and 7.5 million stroke survivors each year [2]. Beside the low control rates of modifiable risk factors, discrepancies in genetic background have been proposed as a potential cause for the uneven distribution of stroke risk in geographic regions and among ethnic populations [3], but the specific genes responsible for the extra risk have not been identified to date.

In the past years, genome-wide association studies (GWAS) have identified several ischemic stroke susceptibility loci, including 4q25, 16q22, 12p13, 7p21, and 6p21 [4–8]. A recent GWAS [9] associated rs505922, a single nucleotide polymorphism (SNP) located in the first intron of ABO gene, with stroke of large artery atherosclerotic (LAA) and

Hao Zhang and Zhizhong Zhang contributed equally to this work.

✉ Gelin Xu
gelinxu@nju.edu.cn

- ¹ Department of Neurology, Jinling Hospital, Medical School of Nanjing University, 305 East Zhongshan Road, Nanjing 210002, Jiangsu, China
- ² Department of Neurology, The First People's Hospital of Hangzhou, Nanjing Medical University, Hangzhou 310006, Zhejiang, China
- ³ Department of Neurology, Jinling Hospital, Southern Medical University, Nanjing 210002, Jiangsu, China
- ⁴ Department of Gerontology, Drum Tower Hospital, Medical School of Nanjing University, Nanjing 210008, Jiangsu, China
- ⁵ Department of Neurology, Huai'an First People's Hospital, Nanjing Medical University, Huai'an 223300, Jiangsu, China

cardioembolic types, but not with stroke of small vessel disease type. This association between ABO gene polymorphisms and risk of specified stroke subtypes were detected only in a Caucasian population and remained to be tested in other ethnicities. Increasing evidence suggests that genetic risks differ depending on ischemic stroke subtype [10]. LAA stroke is the most common subtype in China [11].

Therefore, to validate the findings of the previous Caucasian GWAS study in Chinese Han population and to search for more susceptible loci of LAA stroke, a case–control study consisting of 644 cases and 642 controls was performed.

Material and Methods

Ethics Statement

This study was approved by the Ethical Review Board of Jinling Hospital (Nanjing, China). Written informed consent was obtained from each enrolled participants.

Study Population

Patients were enrolled from Nanjing Stroke Registry Program (NSRP), which has been described in previous publications [12]. Between March 2012 and December 2014, consecutive patients with first-ever stroke were assessed for eligibility of participation. Inclusion criteria of the cases included Chinese Han ethnicity; LAA stroke classified according to TOAST and confirmed by imaging results within 14 days of stroke onset; and aged 18 years or older. Exclusion criteria included malignancies; severe heart, lung, liver, and kidney dysfunction; autoimmune and inflammatory diseases; and hematological diseases.

The controls were screened from local residents who had a regular physical examination. Residents were included if they were ethnically Chinese Han, aged 18 years or older, without history of atherosclerotic diseases, and without history of stroke or cardiovascular diseases.

Fine-Mapping Strategy

By searching the HapMap release 27 of the merged phases 2 and 3 (<http://hapmap.ncbi.nlm.nih.gov/>), a LD block of 25.4 kb around the rs505922 SNP in Han Chinese was identified. Using Haploview software version 4.2 (<http://www.broadinstitute.org/haploview/haploview>), tag SNPs for fine-mapping were determined based on their possibility to tag surrounding variants [13].

The tag SNPs (Table 1) were selected based on the following criteria: (1) with $r^2 = 0.8$; (2) with minor allele frequency (MAF) $\geq 10\%$; (3) with Hardy-Weinberg equilibrium test P value ≥ 0.05 ; and (4) with overall rate of genotype $\geq 75\%$. The index SNP, rs505922, was also included. As a result, ten tag

Table 1 Information of ten selected SNPs of ABO gene region in a Chinese population

SNP	Chr	Location on Chr	Allele	MAF (CHB) ^a	P value ^b
rs630014	9	135139543	C:T	0.433	0.405
rs505922	9	135139050	T:C	0.389	0.750
rs500499	9	135138469	G:C	0.478	0.187
rs8176668	9	135133880	T:A	0.261	0.123
rs575259	9	135125573	A:G	0.311	0.526
rs8176722	9	135122575	G:T	0.216	0.778
rs2073824	9	135122454	G:A	0.341	0.080
rs8176725	9	135122438	C:T	0.326	0.071
rs8176740	9	135121293	T:A	0.411	0.921
rs8176731	9	135122171	G:A	0.432	0.323

Chr chromosome, *MAF* minor allele frequency, *SNP* single nucleotide polymorphism

^a From both HapMap and dbSNPs databases

^b P value for Hardy-Weinberg equilibrium in controls

SNPs (rs630014, rs505922, rs500499, rs8176668, rs575259, rs8176722, rs2073824, rs8176725, rs8176731, and rs8176740) were selected for genotyping. Figure 1a shows rs number and relative position of these ten tSNPs. Figure 1b shows their LD blocks in controls which were calculated by the Haploview 4.2 software.

Blood Sample Collection and Genotyping

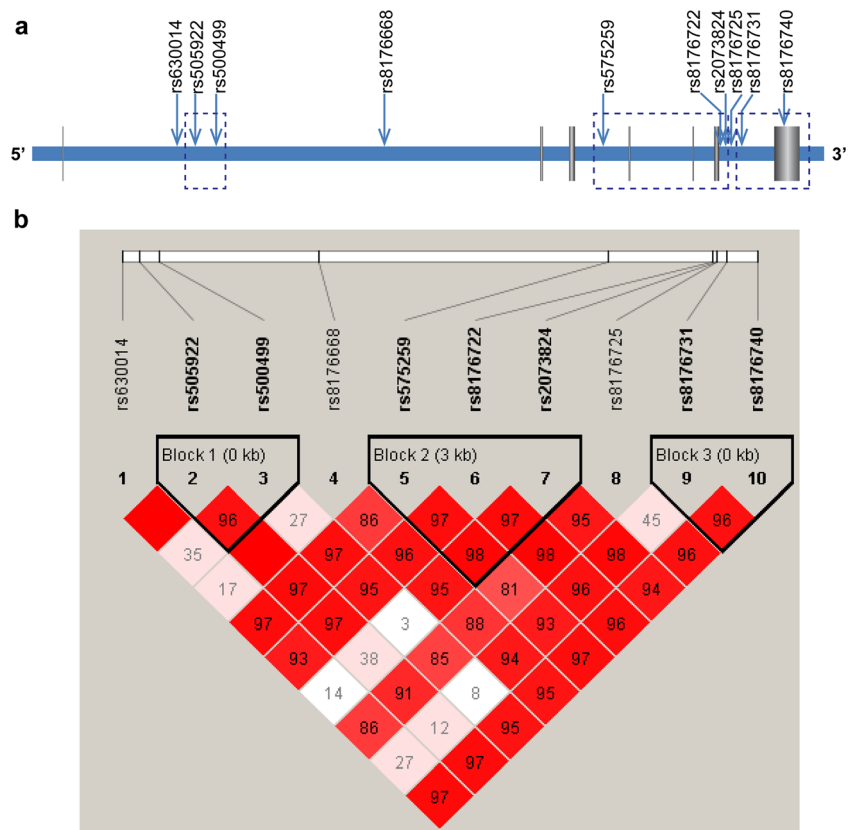
Genomic DNA was extracted with DNA isolation kit. Genotyping for variants of ABO gene was conducted by the Improved Multiple Ligase Detection Reaction (iMLDR) [14], and 5 % randomly selected samples were used to assess reproducibility of genotypes, which resulted in a 100 % concordance.

SNP–SNP Interactions

Multifactor dimensionality reduction (MDR) (version 3.0_0_2) and MDR-permutation testing (MDRpt, version 1.0_beta_2) were performed to evaluate the SNP–SNP interactions in the risk of LAA stroke. In MDR, multilocus genotypes were classified as high- and low-risk groups. With this method, the multidimensional genotype variables were transformed into single-dimensional ones [15]. The validity of these transformed single-dimensional multilocus genotype variables in predicting LAA stroke risk were evaluated with cross-validation and permutation testing [16].

Among multilocus models, the best candidate interaction model was the one with the highest testing accuracy and the cross-validation consistency (CVC). The true positive models would have estimated testing accuracy higher than 0.5. We performed MDR in a cross-validation framework that can

Fig. 1 **a** ABO genomic organization. Exons are presented as *gray blocks*; linkage disequilibrium blocks are indicated by *dotted box*. The SNPs analyzed in the present study are indicated by *arrows* according to their genomic position. **b** LD plot of ABO gene using 10 SNPs in 642 ethnic Han Chinese controls. This plot was generated by the Haploview program. Three blocks were determined. The rs number (*top; from left to right*) corresponds to the SNP name and the level of pairwise. D' indicates the degree of LD between the two SNPs



assess the predictive ability of the models [17]. We implemented 1000-fold permutation testing methods to assess statistical significance. The model was considered to be statistically significant when the P value was less than 0.05.

Statistical Analysis

Power analyses were conducted by PS software, version 3.0.14 (available at <http://www.mc.vanderbilt.edu/prevmed/ps>). A fixed MAF of 30 %, a type I error probability of 0.05, and an odds ratio (OR) of 1.4 were used to estimate the power. The sample of this study yielded a power of 81.2 % in analyzing association between the cases and the controls.

Hardy-Weinberg equilibrium was tested with chi-squared goodness-of-fit test to examine the frequency distribution of each SNP in control subjects. Haploview v4.2 was adopted to analyze linkage disequilibrium (LD) and identify haplotype blocks.

Chi-squared tests were used to compare differences of categorical variables between groups. The Student's t test was performed to compare the continuous covariates. Logistic regression model were used to evaluate association between SNP polymorphisms and risk of LAA stroke. The Hosmer–Lemeshow goodness-of-fit test was performed to assess calibration of logistic regression model. A P value of more than 0.05 was deemed as satisfactory. The discriminative power of

multivariate logistic regression analysis was evaluated with a receiver operating characteristic (ROC) curve.

Haplotypes were reconstructed with PHASE software (version 2.1). In haplotype analysis, haplotype frequencies less than 0.01 were omitted. MDR was adopted to identify the potential interactions of these ten tSNPs on risk of LAA stroke. Bilateral P value less than 0.05 was deemed statistically significant. Statistical analyses were performed using IBM SPSS Statistics version 22.0 (Armonk, NY: IBM Corp.).

Results

Baseline Characteristics

The clinical and demographic characteristics of the cases and controls are shown in Table 2. The cases and controls were well matched on age ($P=0.877$) and sex ($P=0.683$). Prevalence of hyperlipidemia was also much similar between cases and controls ($P=0.767$). As shown in Table 2, the cases had a higher prevalence of conventional risk factors for vascular diseases, including a history of hypertension, diabetes mellitus (DM) and smoking compared with the controls, and body mass index (BMI) was also significantly different between the cases and the controls ($P<0.01$). Thus, these variables (age, sex, BMI, hypertension, DM, hyperlipidemia, and

Table 2 Characteristics of cases and controls

Variables	Cases (<i>n</i> = 644)	Controls (<i>n</i> = 642)	<i>P</i> value
Age (years)			0.877
≥70	75 (11.6 %)	73 (11.4 %)	
<70	569 (88.4 %)	569 (88.6 %)	
BMI (kg/m ²)	24.82 ± 3.02	24.20 ± 2.84	<0.001
Sex			0.683
Male	470 (73.0 %)	462 (72.0 %)	
Female	174 (27.0 %)	180 (28.0 %)	
Hypertension			<0.001
No	215 (33.4 %)	443 (69.0 %)	
Yes	429 (66.6 %)	199 (31.0 %)	
Diabetes			<0.001
No	464 (72.0 %)	587 (91.4 %)	
Yes	180 (28.0 %)	55 (8.6 %)	
Hyperlipidemia			0.767
No	499 (77.5 %)	493 (76.8 %)	
Yes	145 (22.5 %)	149 (23.2 %)	
Smoking			0.042
No	345 (53.6 %)	380 (59.2 %)	
Yes	299 (46.4 %)	262 (40.8 %)	

smoking) were adjusted in the multivariate logistic regression analysis to evaluate the main effects of ABO gene polymorphisms on risk of LAA stroke.

Association Between ABO SNPs and LAA Risk

The genotype distributions in the controls were consistent with Hardy–Weinberg equilibrium (Table 1). In the univariate analyses, the genotype frequencies of two SNPs were significantly different between the cases and the controls: rs8176668 ($P=0.005$) and rs2073824 ($P=0.001$, Table 3). For rs8176668 A>T polymorphism, the frequencies of AA, AT, and TT genotypes were 52.3, 39.1, and 8.6 %, respectively, in cases, and were 43.3, 47.0, and 9.7 %, respectively, in controls. For rs2073824 G>A polymorphism, the frequencies of GG, GA, and AA genotypes were 38.0, 44.6, and 8.8 % in cases and were 28.0, 52.5, and 19.5 % in controls. In multivariate logistic regression analysis, adjusted for age, sex, BMI, hypertension, DM, hyperlipidemia, and smoking, AT genotype of rs8176668 had a significant protective effect on LAA stroke (OR = 0.71; 95 % CI, 0.55 to 0.92). AA genotype of rs2073824 also had protective effects on LAA stroke (OR = 0.65, 95 % CI, 0.46 to 0.93). In the dominant-effect model, both rs8176668 (TT/AT vs AA, OR = 0.721; 95 % CI, 0.57 to 0.92; $P=0.001$) and rs2073824 (AA/GA vs GG, OR = 0.61; 95 % CI, 0.47 to 0.79; $P<0.001$) were associated with LAA stroke. The goodness-of-fit test demonstrated good calibration (for rs8176668, $P=0.553$; for rs2073824,

$P=0.830$). For rs8176668 multivariate logistic regression analysis model, the area below the ROC curve was 0.737 with a 95 % CI of 0.71–0.76. For rs2073824 multivariate logistic regression analysis model, the area below the ROC curve was 0.742 with a 95 % CI of 0.72–0.77.

Haplotype Block Structure and LD Analysis

As shown in Fig. 1, LD analysis showed that there were three blocks in the ABO gene. Seven of ten selected SNPs were included by these three blocks. Rs505922 and rs500499 were located in LD block 1; rs575259, rs8176722, and rs2073824 were in LD block 2; and rs8176731 and rs8176740 were in LD block 3. In haplotype analysis, after adjusting for age, sex, BMI, hypertension, DM, hyperlipidemia, and smoking, haplotype TC (OR = 0.72; 95 % CI, 0.54 to 0.95) and haplotype ACA (OR = 0.73; 95 % CI, 0.58 to 0.91) were associated with a decreased risk of LAA stroke (Table 4).

SNP–SNP Interactions

A total of ten tSNPs in ABO gene were analyzed by MDR. Table 5 summarizes the results of MDR for analyzing interactions of the ten tag SNPs in their influences on risk of LAA stroke. In the single-locus model, rs2073824 was the most influential attributor for LAA stroke risk, with testing accuracy (TA) of 0.55, cross-validation consistency (CVC) of 10/10, which is consistent with the results of previous logistic analysis. In the multilocus model, the best interaction model was a single-locus model incorporating rs2073824 SNP (testing accuracy = 0.5469, CVC = 5/10, $P<0.05$).

Discussion

This study found that rs8176668 and rs2073824 in ABO gene were associated with reduced risk of LAA stroke in Chinese population. The sample size provided 81.2 % power to detect an OR of 1.4 for minor allele frequencies of 0.3. These results are novel to the results of the recent Caucasian GWAS [9]. These findings suggested that susceptibility loci for LAA stroke risk identified by Caucasian GWAS might not play a role in LAA stroke in Chinese population.

These discrepant results between different studies may be explained from two aspects. The first explanation lies in the genetic differences between Chinese and Caucasian populations. The ABO gene varies in geographic/ethnic distribution of allele frequencies [18]. A previous study identified three novel ABO alleles in the Chinese Han population [19]. Differences of minor allele frequencies between Chinese and Caucasian populations were also observed. The minor allele in CEU (Utah residents with Northern and Western European ancestry from the CEPH collection) become the major allele

Table 3 Association between tag SNPs and risk of LAA stroke

Genotypes	Cases <i>n</i> = 644	Controls <i>n</i> = 642	<i>P</i> value	Crude OR (95 % CI)	Adjusted OR (95 % CI) ^a
rs505922					
CC	173 (26.9 %)	191 (29.8 %)	0.377	1.00	1.00
CT	325 (50.5 %)	322 (50.2 %)		1.11(0.86–1.44)	1.20 (0.91–1.60)
TT	146 (22.6 %)	129 (20.0 %)		1.25 (0.91–1.71)	1.32 (0.94–1.87)
TT+CT	471 (73.1 %)	451 (70.2 %)	0.251	1.15 (0.90–1.47)	1.24 (0.95–1.62)
rs575259					
AA	364 (56.5 %)	389 (60.6 %)	0.204	1.00	1.00
GA	241 (37.4 %)	225 (35.0 %)		1.15 (0.91–1.44)	1.08 (0.84–1.39)
GG	39 (6.1 %)	28 (4.4 %)		1.49 (0.90–2.47)	1.25 (0.72–2.19)
GG+GA	280 (43.5 %)	253 (39.4 %)	0.138	1.18 (0.95–1.48)	1.10 (0.86–1.40)
rs8176722					
CC	379 (58.9 %)	373 (58.1 %)	0.916	1.00	1.00
CA	225 (34.9 %)	231 (36.0 %)		0.96 (0.76–1.21)	1.05 (0.82–1.36)
AA	40 (6.2 %)	38 (5.9 %)		1.04 (0.65–1.65)	1.11 (0.67–1.83)
AA+CA	265 (41.1 %)	269 (41.9 %)	0.785	0.97 (0.78–1.21)	1.06 (0.83–1.36)
rs8176668					
AA	337 (52.3 %)	278 (43.3 %)	0.005	1.00	1.00
AT	252 (39.1 %)	302 (47.0 %)		0.69 (0.55–0.87)	0.71 (0.55–0.92)
TT	55 (8.6 %)	62 (9.7 %)		0.73 (0.49–1.09)	0.76 (0.49–1.18)
TT+AT	307 (47.7 %)	364 (56.7 %)	0.001	0.70 (0.56–0.87)	0.72 (0.57–0.92)
rs8176731					
TT	179 (27.8 %)	189 (29.4 %)	0.426	1.00	1.00
CT	323 (50.2 %)	330 (51.4 %)		1.03 (0.80–1.33)	1.01 (0.76–1.34)
CC	142 (22.0 %)	123 (19.2 %)		1.22 (0.89–1.67)	1.21 (0.86–1.71)
CC+CT	465 (72.2 %)	453 (70.6 %)	0.514	1.08 (0.85–1.38)	1.06 (0.82–1.39)
rs500499					
CC	218 (33.9 %)	226 (35.2 %)	0.766	1.00	1.00
GC	310 (46.7 %)	296 (46.1 %)		1.09 (0.85–1.39)	1.04 (0.79–1.36)
GG	116 (19.4 %)	120 (18.7 %)		1.00 (0.73–1.37)	0.93 (0.66–1.32)
GG+GC	426 (66.1 %)	416 (64.9 %)	0.514	1.06 (0.84–1.34)	1.01 (0.78–1.30)
rs8176740					
AA	384 (59.6 %)	403 (62.8 %)	0.340	1.00	1.00
AT	226 (35.1 %)	214 (33.3 %)		1.11 (0.88–1.40)	1.07 (0.83–1.38)
TT	34 (5.3 %)	25 (3.9 %)		1.43 (0.84–2.44)	1.27 (0.71–2.30)
TT+AT	260 (40.6 %)	239 (37.2 %)	0.247	1.14 (0.91–1.43)	1.09 (0.85–1.40)
rs8176725					
GG	258 (40.1 %)	278 (43.2 %)	0.090	1.00	1.00
GA	282 (43.8 %)	287 (44.7 %)		1.06 (0.84–1.34)	1.15 (0.89–1.50)
AA	104 (16.1 %)	77 (12.1 %)		1.46 (1.04–2.04)	1.60 (1.10–2.31)
AA+GA	386 (59.9 %)	364 (56.8 %)	0.239	1.14 (0.92–1.43)	1.25 (0.98–1.60)
rs2073824					
GG	245 (38.0 %)	180 (28.0 %)	0.001	1.00	1.00
GA	287 (44.6 %)	337 (52.5 %)		0.63 (0.49–0.80)	0.60 (0.45–0.78)
AA	112 (17.4 %)	125 (19.5 %)		0.66 (0.48–0.90)	0.65 (0.46–0.93)
AA+GA	399 (72.0 %)	462 (72.0 %)	1.741×10^{-4}	0.64 (0.50–0.80)	0.61 (0.47–0.79)
rs630014					
GG	260 (40.4 %)	254 (39.6 %)	0.834	1.00	1.00
GA	306 (47.5 %)	315 (49.1 %)		0.95 (0.75–1.20)	0.95 (0.74–1.23)
AA	78 (12.1 %)	73 (11.3 %)		1.04 (0.73–1.50)	0.91 (0.61–1.36)
AA+GA	384 (59.6 %)	388 (88.7 %)	0.767	0.97 (0.77–1.21)	0.95 (0.74–1.20)

CI confidence interval, OR odds ratio

^a Adjusted for age, sex, body mass index, hypertension, diabetes mellitus, hyperlipidemia, and smoking

Table 4 Haplotype analysis between cases and controls

HapMap block	Haplotype ^a	Case (<i>n</i> , frequencies)	Control (<i>n</i> , frequencies)	Crude OR (95 % CI)	<i>P</i> value	Adjusted OR (95 % CI) ^b	<i>P</i> value ^b
Block 1	CC	611, 0.4739	572, 0.4445	1.0		1.0	
	TG	536, 0.4157	528, 0.4102	0.95 (0.81–1.12)	0.547	0.97 (0.81–1.17)	0.759
	TC	135, 0.1053	176, 0.1381	0.72 (0.56–0.92)	0.010	0.72 (0.54–0.95)	0.018
Block 2	GCG	302, 0.2324	272, 0.2124	1.0		1.0	
	ACA	496, 0.3853	584, 0.4524	0.77 (0.72–0.94)	0.010	0.73 (0.58–0.91)	0.005
	ACG	170, 0.1341	120, 0.0943	1.28 (0.96–1.70)	0.094	1.36 (0.99–1.86)	0.054
	AAG	303, 0.2329	297, 0.2315	0.92 (0.73–1.16)	0.469	0.88 (0.69–1.13)	0.327
Block 3	TA	678, 0.5255	704, 0.5438	1.0		1.0	
	CA	316, 0.2462	316, 0.2506	1.04 (0.86–1.25)	0.695	1.02 (0.83–1.25)	0.876
	CT	291, 0.2250	260, 0.1980	1.16 (0.95–1.42)	0.136	1.22 (0.98–1.51)	0.078

CI confidence interval, *OR* odds ratio

^a Haplotypes with frequency less than 1 % were omitted

^b Adjustment for age, sex, body mass index, hypertension, diabetes, hyperlipidemia, and smoking

in CHB (Han Chinese in Beijing, China) (rs2073824, G allele, CEU = 0.372, CHB = 0.608). The second explanation lies in that rs505922 may be a marker polymorphism rather than a causal variant.

The SNP rs505922 detected in Caucasian GWAS is located in the first intron of ABO gene, which was associated with LAA stroke. This study identified two novel SNPs, rs8176668 and rs2073824, associated with LAA stroke. Rs8176668 and rs2073824 were located in the first and the sixth introns of ABO gene, respectively. Two SNPs identified in this study were novel to the one detected in Caucasian GWAS study, but they all located in the ABO gene. ABO gene is located on the distal long arm of chromosome 9 (9q34.1–q3.2) [20] and consists of seven exons and six introns spanning about 18–20 kb. ABO gene encodes a glycosyltransferase, which

transfer monosaccharides to the cell-surface H antigens, and forms the antigenic structure of the ABO blood groups [21]. Several studies [22, 23] have showed that the ABO gene was involved in the development of coronary artery disease, thrombotic events, and hemorrhagic stroke. While introns do not encode protein, they play essential roles in mRNA process and transportation. Some introns are known to enhance the expression of genes which are involved in a process named as intron-mediated enhancement. We speculate that SNPs rs8176668 and rs2073824 may influence the nucleotide splicing process, result in various spliceosomes and reduce the protein product of ABO gene.

MDR is a powerful method for analyzing gene–gene interactions and has been widely used in several genetic association studies of complex diseases. So, this method is suitable

Table 5 SNP–SNP interactions analyzed with MDR

Model	Bal.Acc.Cv training	Test.Acc.Cv training	CVC	<i>P</i> value ^a
rs2073824	0.5500	0.5500	10/10	0.022
rs2073824, rs575259	0.5562	0.5469	5/10	0.260
rs8176725, rs575259, rs8176668	0.5648	0.5134	4/10	0.542
rs8176725, rs575259, rs8176668, rs505922	0.573	0.5126	3/10	0.559
rs8176740, rs8176725, rs575259, rs8176668, rs500499	0.5796	0.5118	3/10	0.575
rs8176740, rs8176725, rs575259, rs8176668, rs500499, rs505922	0.5836	0.5002	3/10	0.787

CVC cross-validation consistency

^a *P* value based on 1000 permutations

for exploring the relationship between SNP–SNP interactions and risk of LAA stroke. However, no significant SNP–SNP interaction among these ten SNPs was observed.

This study has limitations that need to be considered when interpreting the results. First, the cases and controls were recruited from hospitals, which might result in potential selection bias. Second, we provided evidence for association between ABO gene polymorphisms and risk of LAA stroke, but the underlying molecular mechanisms are still unknown, and need to be determined by cell or animal studies in future. Furthermore, we did not exclude residents with family history of cardiovascular disease or stroke in control selection, which might introduce extra bias. The controls with family history of cardiovascular disease may have higher frequencies of genotypes associated with cardiovascular disease. Therefore, exclude those controls with family history of cardiovascular disease may decrease the likelihood of type II error (false negative).

In conclusion, rs8176668 and rs2073824 in ABO gene may be associated with LAA stroke in the Chinese Han population, which may be used as candidate biomarkers for LAA stroke susceptibility screening. However, the molecular mechanism by which the ABO gene influenced LAA stroke risk warrant further study.

Acknowledgments This study was supported by the National Natural Science Foundation of China (81501027, 81571148, 81501193, and 81220108008) and Hangzhou Science and Technology Bureau Program (20150733Q11). Authors are grateful to all participants for their cooperation in this study. We also thank Miss Ying Lin and Miss Zhaojun Wang in Jinling Hospital, Medical School of Nanjing University, for their technical supports.

Compliance and Ethical Standard

Conflict of Interest The authors declare that they have no conflicts of interest.

References

- Krishnamurthi RV, Feigin VL, Forouzanfar MH et al (2013) Global and regional burden of first-ever ischaemic and haemorrhagic stroke during 1990–2010: findings from the Global Burden of Disease Study 2010. *Lancet Global Health* 1(5):e259–e281
- Johnston SC, Mendis S, Mathers CD (2009) Global variation in stroke burden and mortality: estimates from monitoring, surveillance, and modelling. *Lancet Neurol* 8(4):345–354
- Humphries SE, Morgan L (2004) Genetic risk factors for stroke and carotid atherosclerosis: insights into pathophysiology from candidate gene approaches. *Lancet Neurol* 3(4):227–235
- Gretarsdottir S, Thorleifsson G, Manolescu A et al (2008) Risk variants for atrial fibrillation on chromosome 4q25 associate with ischemic stroke. *Ann Neurol* 64(4):402–409
- Gudbjartsson DF, Holm H, Gretarsdottir S et al (2009) A sequence variant in ZFX3 on 16q22 associates with atrial fibrillation and ischemic stroke. *Nat Genet* 41(8):876–878
- Ikram MA, Seshadri S, Bis JC et al (2009) Genomewide association studies of stroke. *N Engl J Med* 360(17):1718–1728
- International Stroke Genetics C, Wellcome Trust Case Control C, Bellenguez C et al (2012) Genome-wide association study identifies a variant in HDAC9 associated with large vessel ischemic stroke. *Nat Genet* 44(3):328–333
- Holliday EG, Maguire JM, Evans TJ et al (2012) Common variants at 6p21.1 are associated with large artery atherosclerotic stroke. *Nat Genet* 44(10):1147–1151
- Williams FM, Carter AM, Hysi PG et al (2013) Ischemic stroke is associated with the ABO locus: the EuroCLOT study. *Ann Neurol* 73(1):16–31
- Munshi A, Das S, Kaul S (2015) Genetic determinants in ischaemic stroke subtypes: seven year findings and a review. *Gene* 555(2): 250–259
- Liu M, Wu B, Wang WZ et al (2007) Stroke in China: epidemiology, prevention, and management strategies. *Lancet Neurol* 6(5): 456–464
- Liu X, Xu G, Wu W et al (2006) Subtypes and one-year survival of first-ever stroke in Chinese patients: The Nanjing Stroke Registry. *Cerebrovasc Dis* 22(2–3):130–136
- Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21(2): 263–265
- Thomas G, Sinville R, Sutton S et al (2004) Capillary and micro-electrophoretic separations of ligase detection reaction products produced from low-abundant point mutations in genomic DNA. *Electrophoresis* 25(10–11):1668–1677
- Ritchie MD, Hahn LW, Roodi N et al (2001) Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. *Am J Hum Genet* 69(1):138–147
- Sharma KL, Rai R, Srivastava A et al (2014) A multigenic approach to evaluate genetic variants of PLCE1, LXRs, MMPs, TIMP, and CYP genes in gallbladder cancer predisposition. *Tumour Biology* 35(9):8597–8606
- Coffey CS, Hebert PR, Ritchie MD et al (2004) An application of conditional logistic regression and multifactor dimensionality reduction for detecting gene-gene interactions on risk of myocardial infarction: the importance of model validation. *BMC bioinformatics* 5:49
- Chester MA, Olsson ML (2001) The ABO blood group gene: a locus of considerable genetic diversity. *Transfus Med Rev* 15(3): 177–200
- Zhu F, Tao S, Xu X et al (2010) Distribution of ABO blood group allele and identification of three novel alleles in the Chinese Han population. *Vox Sang* 98(4):554–559
- Bennett EP, Steffensen R, Clausen H, Weghuis DO, Geurts van Kessel A (1995) Genomic cloning of the human histo-blood group ABO locus. *Biochem Biophys Res Commun* 211(1):347
- Reid ME, Mohandas N (2004) Red blood cell blood group antigens: structure and function. *Semin Hematol* 41(2):93–117
- Schunkert H, König IR, Kathiresan S et al (2011) Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat Genet* 43(4):333–338
- Wiggins KL, Smith NL, Glazer NL et al (2009) ABO genotype and risk of thrombotic events and hemorrhagic stroke. *J Thromb Haemost* 7(2):263–269