Identification of the susceptibility gene loci associated with ischemic stroke in a Mongolian population in China

Qizhu Wu*, Huiguang Wu*, Gegentana Gegentana*, Wenyan Huo, Suyalatu Suyalatu, Nuomin Wu, Ming Chen, Haihua Bai#

Inner Mongolia University for the Nationalities, Tongliao 028000, China, hh_bai@163.com

**These authors contributed equally to this work. # Correspondence author.*

Ischemic stroke, accounting for 87% of the stroke population [1], is a leading cause of death and disability worldwide, and presents a serious and growing threat to public health. More than 40 common sequence variants associated with ischemic stroke had been identified by genome-wide association studies (GWAS). However, the discovery of new susceptibility loci in different ethnic groups emphasizes the need for conducting more GWAS in populations of diverse ethnic groups. To provide new insights in clinical diagnosis and therapy, we analyzed the susceptibility gene loci associated with ischemic stroke in a Mongolian population.

The subjects included 343 individuals of Chinese Mongolian ethnicity, of whom 167 were patients affected with ischemic stroke at various stages of the disease and 176 were normal controls. Control individuals were not related to the ischemic stroke patients or each other. Aside from the diagnosis of ischemic stroke, we collected the related clinical characteristics, such as the body mass index (BMI), systolic blood pressure (SBP), diastolic blood pressure (DBP), total cholesterol (TC), triglyceride (TG), high density lipoprotein cholesterol (HDL-C) and low density lipoprotein cholesterol (LDL-C). A multivariate unconditional logistic regression model was employed to analyze the high risk factors of ischemic stroke. We selected 4 risk factors for ischemic stroke, which were sex, age, BMI, high blood pressure (HBP).

We selected 15 single nucleotide polymorphisms (SNPs) sites located in or near 15 candidate genes (Table 1) of ischemic stroke based on NHGRI GWAS catalog (www.genome.gov/gwastudies, November, 2012). We amplified the 15 SNP loci using Techne Prime Thermal Cycler. After the PCR amplifications, we purified the PCR products using 1X Agencourt AMPure XP-Medium beads and acquired the mixed Illumina pair-end

libraries. The insert sizes of libraries were calculated by Aglient 2100 bio-analyzer. The concentrations of libraries were estimated by Real Time PCR. The samples were sequencing using the Illumina MiSeq. All sequencing steps were in strict accordance with Illumina recommended protocols.

The sequencing depths exceed 200X for each locus, and the length of pair-end reads is 100bp. The reads with too many N bases (>10%) or low base quality (>50% bases with base quality < 5) were discarded.

We mapped all clean reads against the human reference genome (hg19) using the BWA [2] (version 0.5.9) with the cutoff of 3bp mismatches across a single read. Samtools [3] (version 0.1.18) was used to call to obtain raw SNP genotypes. We adopted the following criteria to filtering the raw SNP genotypes: SNPs with >=5% of missing call rate across the samples and samples with >=3% of missing genotypes were removed. We tested the genotypes of the SNPs for Hardy-Weinberg disequilibrium, and dropped the ones that have P $<1x10^{-6}$ in unaffected individuals. All 15 SNPs loci passed through these quality control.

We tested association between candidate gene SNPs and ischemic stroke using logistic regression. Estimated odds ratios (ORs) are equivalent to those obtainable from the Cochran–Armitage trend test. Under the widely accepted additive genetic model for this disorder, the trend test is more robust to deviations from Hardy–Weinberg equilibrium and is hence preferred to other tests, such as those calculated by contrasting allele frequencies or homozygous frequencies. The threshold *P*-value for association was set at 0.05 by applying Bonferroni correction for testing 15 SNPs. Formal statistical tests and parameter estimations, including 95% confidence intervals (CI), were performed using PLINK [4] (version 1.07). Multivariate logistic regression analysis was performed by using SPSS (version 18.0).

The results showed that rs12425791 of *NINJ2* (OR=1.83, *P*=0.0044), rs2238151 of *ALDH2* (OR=1.74, *P*=0.0208) and rs9536591 of *intergenic* (OR=1.52, *P*=0.0162) were significantly associated with ischemic stroke risks (Table 1). *CDC5L* and *HDAC9* is risk factors of ischemic stroke in other group [5, 6]. We found there is no association between *CDC5L* (OR=0.59, *P*=0.0036), *HDAC9* (OR=0.57, *P*=0.0018) and the ischemic stroke in the Chinese Mongolian population.

Table 1 15 SNPs of susceptibility gene loci of ischemic stroke

Chr	Gene	SNP	Minor allele	MAF		OR	95% CI	*P* value (adjusted for age, sex, BMI)
				Cases	Controls			
3	*SPSB4*	rs16851055	A	0.126	0.128	0.84	0.52-1.36	0.478
4	*NR*	rs2200733	T	0.473	0.446	1.06	0.75-1.5	0.7525
4	*PITX2*	rs6843082	A	0.27	0.335	0.76	0.51-1.15	0.198
5	*ADAMTS12*	rs1364044	T	0.422	0.421	1.06	0.74-1.54	0.7406
5	*ADAMTS2*	rs469568	C	0.174	0.114	1.53	0.99-2.36	0.057
6	*CDC5L*	rs556621	T	0.434	0.443	0.59	0.42-0.84	0.0036
6	*AIM1*	rs783396	A	0.054	0.08	0.79	0.39-1.57	0.4994
7	*Intergenic*	rs10486776	A	0.057	0.071	0.79	0.41-1.53	0.4824
7	*HDAC9*	rs2107595	A	0.225	0.335	0.57	0.41-0.81	0.0018
11	*TRIM29*	rs2084898	A	0.018	0.026	0.78	0.25-2.46	0.6773
12	*NINJ2*	rs12425791	A	0.374	0.27	1.83	1.21-2.77	0.0044
12	*ALDH2*	rs2238151	T	0.189	0.139	1.74	1.09-2.79	0.0208
13	*Intergenic*	rs9536591	A	0.407	0.315	1.52	1.08-2.15	0.0162
16	*ZFHX3*	rs879324	A	0.293	0.341	0.72	0.48-1.07	0.0991
18	*IMPA2*	rs7506045	T	0.27	0.21	1.22	0.85-1.77	0.283

Note: Chr, chromosome; MAF, minor allele frequency; CI, confidence interval. Adjusted *P* values, odds ratios, and 95% confidence intervals were calculated using the additive model of genetic association.

In conclusion, we identified SNPs rs12425791 of *NINJ2*, rs2238151 of *ALDH2*, rs9536591 of *intergenic* associated with ischemic stroke in the Chinese Mongolian population. Our discovery also demonstrated genetic heterogeneity exists between Chinese Mongolian population versus other populations.

We thank our study individuals for their generous participation in this study. This study was supported by the National Natural Science Foundation of China (81060098, 81160101).

References

1. Rosamond W, et al. (2008) . American Heart Association Statistics Committee and Stroke Statistics Subcommittee Heart disease and stroke statistics—2008 update: a report from the American Heart Associations Statistics Committee and Stroke Statistics Subcommittee, *Circulation*, **117:** e25–e146.

2. Li H., Durbin R. (2009) Fast and accurate short read alignment with Burrows-Wheeler Transform, *Bioinformatics*, **25:** 1754-60.

3. Li H. et al. (2009) The Sequence alignment/map (SAM) format and SAMtools, *Bioinformatics*, **25:** 2078-9.

4. Purcell S, et al. (2007) PLINK: a toolset for whole-genome association and population-based linkage analysis, *American Journal of Human Genetics*, **81:** 559–575.

5. Holliday E G, et al. (2012) Common variants at 6p21.1 are associated with large artery atherosclerotic stroke, *Nature genetics*, 44: 1147-1151.

6. Bellenguez C, et al. (2012) Genome-wide association study identifies a variant in HDAC9 associated with large vessel ischemic stroke, *Nature genetics*, 44: 328-333.