Original Article

Genetic variant in ZFHX3 Gene on 16q22 associated with risk of stroke in Chinese Han population

Yidong Xue1*, Ning Shi1*, Bolun Zhang2, Xiaorong Gao1, Yajun Gao1, Xuan Zhou1, Jianfeng Du1, Peng Chen3,4, Yongri Ouyang4, Fengjiao Wang4, Tianbo Jin4

1Department of Neurology, The Affiliated Hospital of Yan’an University, Yan’an, Shaanxi, China; 2Second Affiliated Hospital of Inner Mongolia Medical University, Hohhot, China; 3Institution of Basic Medical Science, Xi’an Medical University, Xi’an 710021, Shaanxi, China; 4National Engineering Research Center for Miniaturized Detection Systems, School of Life Sciences, Northwest University, Xi’an, China. *Co-first authors.

Received January 25, 2016; Accepted June 13, 2016; Epub August 1, 2016; Published August 15, 2016

Abstract: In China, Stroke is one of the most prevalent cerebrovascular disease and one of the leading causes of death and chronic disability in men. Studies have identified that stroke is a complex and heterogeneous disease caused by both genetic and environmental factors. A genome-wide association study (GWAS) in patients of European population has identified that risk variants in ZFHX3 on 16q22 was related with stroke and atrial fibrillation (AF) while AF increases the risk of stroke 4-5 fold across all age groups and accounts for 10-15% of all ischemic stroke. However, this association has not been tested in Chinese Han Population. We aim to figure out the sequence variants if it increased the risk of stroke in Chinese Han descent. Six SNPs of ZFHX3 gene were genotyped, and association analysis was performed. Odds ratios (ORs) and 95% confidence intervals (CIs) were calculated in logistic regression models. We analyzed a Chinese Han cohort consisting of 488 stroke patients and 503 non-stroke controls to test whether the previous findings on ZFHX3 have significant association on stroke with a different ethnic population. We cannot find any significant association for rs16971384, rs12596992, rs17680796, rs17681554 and rs879324 in ZFHX3. However, the significant association between rs13336412 and rs9921395 in ZFHX3 and stroke was identified in the dominant model (OR = 0.7301, 95% CI: 0.558-0.955 P = 0.02; OR = 0.723, 95% CI: 0.526-0.994 P = 0.046), indicating it as a prostate stroke risk variant in populations from China. The present study suggests that the ZFHX3 genes polymorphisms are not genetic markers of susceptibility to stroke in Chinese Han patients. The study expands the association between ZFHX3 and AF to a non-European ancestry population and provides the first evidence of a cross-race susceptibility of the 16q22 AF locus.

Keywords: Single nucleotide polymorphism (SNP), stroke, ZFHX3, atrial fibrillation (AF)

Introduction

Stroke is the secondary common cause of death around the world, the global challenge of stroke is huge with 16 million new cases each year and about 6 million deaths [1]. Especially the rural China population incurs greatest burden where there is a stroke pandemic with increasing stroke incidence and prevalence and aging of [2]. However, stroke carries considerable recurrent risk and most of patients have some degree of disability [3, 4]. In addition, at least 25% of people at middle age (40 year old and older) have atrial fibrillation (AF) [5], the risk of stroke in patients with AF is four to five-fold than the normal [6]. Particularly, AF is a pivotal danger factor for stroke [7]. However, recent studies have identified that single nucleotide polymorphisms (SNPs) are associated with an increased stroke risk in the general population [8]. We want to figure out the sequence variants if it increased the risk of stroke in Chinese Han descent. The ZFHX3 gene was identified as a risk factor for stroke and associated with different stroke subtypes. In our study, the significant association between rs13336412 and rs9921395 in ZFHX3 and stroke was identified in the dominant model in Chinese Han populations. However, further large and functional studies are needed to confirm our findings.

Materials and methods

Subjects

This study was approved by the local Ethics Committees and all study consent for their
participation in the present study. The cases were stroke patients recruited from Department of Neurology, the Affiliated Hospital of Yan’an University. All stroke patients were diagnosed by expert neurologists and based on the standard diagnostic criteria given by clinical doctor. The control subjects were recruited from the health checkup center and all of them visited for an annual health examination. All the controls were confirmed to be free of stroke at the time of enrollment. At last, 488 stroke patients and 503 non-stroke controls were recruited among Chinese Han population. All patient information was adjusted by age & gender.

Demographic and clinical data

Demographic and personal detailed information were collected by nurse, including age, sex, ethnicity, residential region, smoking status, alcohol use and education status. For patients, detailed clinical information was collected through a medical chart review or consultation with treating physicians. At least 5-ml venous blood was collected from each subject.

SNP selection and genotyping

First of all, All the 6 SNPs in ZFHX3 gene were associated with stroke which with minor allele frequencies > 5% in the HapMap Han Chinese population and finally a total of 6 SNPs were selected for future genotyping. Using the Haploview 4.2 software, linkage disequilibrium (LD) analysis with an r² ≥ 0.8 was further applied to filter these SNPs and six SNPs (rs16971384, rs17680796, rs17681554, rs879324, rs13336412 and rs9921395) were remained. GoldMag-Mini Purification Kit (GoldMag Co. Ltd. Xian city, China) method was performed to extract genomic DNA from whole blood. DNA concentration was measured by NanoDrop 2000 (Thermo Scientific, Waltham, Massachusetts, USA). Sequenom MassARRAY Assay Design 3.0 software was used to design multiplexed SNP MassEXTEND assay, and SNP genotyping was performed utilizing the Sequenom MassARRAY RS1000 recommended by the manufacturer [9]. Sequenom Typer 4.0 software was used to perform data management and analyses [9-11].

Statistical analysis

We performed statistical analyses by using Microsoft Excel and SPSS 16.0 (SPSS, Chicago, IL, USA). In this study, all two-sided P ≤ 0.05 was considered as achieving the threshold of statistical significance. A Fisher’s exact test was used to test for deviations from Hardy-Weinberg equilibrium (HWE) in each SNP frequency in the control subjects. Odds ratios (ORs) and 95% confidence intervals (CIs) were determined using unconditional logistic regression analysis by adjusting for age and gender [12]. Allele frequencies and genotype frequencies for each SNP of cases and controls were compared using the Chi-squared test/Fisher’s exact test [13, 14].

We performed statistical analyses by using Microsoft Excel and SPSS 16.0 (SPSS, Chicago, IL, USA). In this study, all two-sided P ≤ 0.05 was considered as achieving the threshold of statistical significance. A Fisher’s exact test was used to test for deviations from Hardy-Weinberg equilibrium (HWE) in each SNP frequency in the control subjects. Odds ratios (ORs) and 95% confidence intervals (CIs) were determined using unconditional logistic regression analysis by adjusting for age and gender [12]. Allele frequencies and genotype frequencies for each SNP of cases and controls were compared using the Chi-squared test/Fisher’s exact test [13, 14].

Three genetic models (dominant, recessive, and additive) were accessed by using PLINK software (http://pngu.mgh.harvard.edu/purcell/plink/) to estimate ORs for SNP main effects. The SNP we are looking at has 2 alleles, A and B. Select allele A to model. Thus, the dominant model is coded as 1 for any genotype that contains an A allele and 0 otherwise (i.e. AB, AA = 1, BB = 0); the recessive model is coded as 1 if they were homozygous for the risk allele, and 0 otherwise (i.e. AA = 1, AB, BB = 0); For the additive model, testing is designed specifically to reveal associations that depend additively upon the minor allele—that is, where having two minor alleles (AA) rather than having no minor alleles (aa) is twice as likely to affect the outcome in a certain direction as is having just one minor allele (Aa) rather than no minor alleles (aa), individuals were coded a 0, 1, or 2, representing the number of risk alleles they possessed for that SNP (i.e. AA = 2, AB = 1, BB = 0). We determined p values for trend by entering the variable as a single term in the model and testing using the Wald’s test. For SNP main effects analysis, we used ordinal variables coded as the number of variant alleles, zero, one, or two, assuming a log-additive genetic model. ORs and 95% CIs were

### Table 1. Characteristics of cases and controls in this study

<table>
<thead>
<tr>
<th>Age and Gender</th>
<th>Case %</th>
<th>Control %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>488</td>
<td>503</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>163</td>
<td>33.40%</td>
</tr>
<tr>
<td>Male</td>
<td>325</td>
<td>66.60%</td>
</tr>
<tr>
<td>Mean ± SD</td>
<td>63.96 ± 11.06</td>
<td>50.36 ± 7.89</td>
</tr>
</tbody>
</table>

We performed statistical analyses by using Microsoft Excel and SPSS 16.0 (SPSS, Chicago, IL, USA). In this study, all two-sided P ≤ 0.05 was considered as achieving the threshold of statistical significance. A Fisher’s exact test was used to test for deviations from Hardy-Weinberg equilibrium (HWE) in each SNP frequency in the control subjects. Odds ratios (ORs) and 95% confidence intervals (CIs) were determined using unconditional logistic regression analysis by adjusting for age and gender [12]. Allele frequencies and genotype frequencies for each SNP of cases and controls were compared using the Chi-squared test/Fisher’s exact test [13, 14].

Three genetic models (dominant, recessive, and additive) were accessed by using PLINK software (http://pngu.mgh.harvard.edu/purcell/plink/) to estimate ORs for SNP main effects. The SNP we are looking at has 2 alleles, A and B. Select allele A to model. Thus, the dominant model is coded as 1 for any genotype that contains an A allele and 0 otherwise (i.e. AB, AA = 1, BB = 0); the recessive model is coded as 1 if they were homozygous for the risk allele, and 0 otherwise (i.e. AA = 1, AB, BB = 0); For the additive model, testing is designed specifically to reveal associations that depend additively upon the minor allele—that is, where having two minor alleles (AA) rather than having no minor alleles (aa) is twice as likely to affect the outcome in a certain direction as is having just one minor allele (Aa) rather than no minor alleles (aa), individuals were coded a 0, 1, or 2, representing the number of risk alleles they possessed for that SNP (i.e. AA = 2, AB = 1, BB = 0). We determined p values for trend by entering the variable as a single term in the model and testing using the Wald’s test. For SNP main effects analysis, we used ordinal variables coded as the number of variant alleles, zero, one, or two, assuming a log-additive genetic model. ORs and 95% CIs were
ZFHX3 polymorphism and stroke

Results

In study, 488 cases and 503 controls were carried out to test whether SNPs in the ZFHX3 are associated with stroke from Chinese Han cohort who lived in Shanxi province or nearby regions. The characteristics of the study population are summarized in Table 1. As shown, the average age for stroke cases and controls was 63.96 ± 11.06 and 50.36 ± 7.89 years, so the differences of distributions of age and gender between cases and controls are not statistically significant. All 6 SNPs were in Hardy-Weinberg equilibrium (P > 0.05) (Table 2) in control subjects. There were no significant allelic association between all SNPs and stroke (Table 2). The differences in frequency of genotype between cases and controls by Chi-squared test were compared but we can’t find any significant SNPs that were associated with stroke risk in the ZFHX3 gene at a 5% level. The relevant information for all these SNPs is available from Table 2. The correlation between all SNPs and MAF under multiple inheritance models was observed when we assumed that MAF of each SNPs was risk factor. Logistic regression analyses revealed that functional SNPs rs13336412 and rs9921395 were associated with decreased risk of stroke in the Dominant model (OR = 0.730, 95% CI: 0.558-0.955, P = 0.02; OR = 0.723, 95% CI: 0.526-0.994, P = 0.046) (Table 3). While adjusted by age and gender, association with risk of stroke in rs13336412 was remained (OR = 0.691, 95% CI: 0.487-0.980, P = 0.04).

We also performed haplotype analysis for associations between stroke and multiple SNPs and found strong linkage of these candidate SNPs. In block, a 1 kb genomic region involving two SNPs rs17680796 and rs13336412. Unfortunately, none of the constructed haplotypes within ZFHX3 genes confers the risk for stroke after adjusted by age and gender.

Discussion

Studies have identified that stroke is a complex and heterogeneous disease caused by both

<table>
<thead>
<tr>
<th>SNP</th>
<th>Sample</th>
<th>Genotype distribution</th>
<th>P*</th>
<th>Allele distribution</th>
<th>OR (95% CI)</th>
<th>P*</th>
<th>HWE p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs16971384</td>
<td>Case</td>
<td>GG 233 GA 187 AA 187</td>
<td>0.828</td>
<td>G 367 607 A 0.96</td>
<td>(0.80-1.15)</td>
<td>0.633</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>70 248 183</td>
<td>0.098</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs17680796</td>
<td>Case</td>
<td>TT 188 TC 259 CC 259</td>
<td>0.395</td>
<td>T 270 706 C 0.94</td>
<td>(0.77-1.14)</td>
<td>0.533</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>38 215 250</td>
<td>0.006</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs13336412</td>
<td>Case</td>
<td>AA 217 AC 173 CC 173</td>
<td>0.066</td>
<td>A 411 563 C 0.85</td>
<td>(0.71-1.02)</td>
<td>0.072</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>106 252 144</td>
<td>0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs9921395</td>
<td>Case</td>
<td>AA 79 AG 405 GG 405</td>
<td>0.109</td>
<td>A 85 889 G 0.76</td>
<td>(0.56-1.02)</td>
<td>0.063</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>3 107 393</td>
<td>0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs7819554</td>
<td>Case</td>
<td>CC 151 CA 324 AA 324</td>
<td>0.014</td>
<td>C 175 799 A 0.85</td>
<td>(0.68-1.07)</td>
<td>0.16</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>19 167 315</td>
<td>0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs879324</td>
<td>Case</td>
<td>AA 219 AG 223 GG 223</td>
<td>0.194</td>
<td>A 311 665 G 0.86</td>
<td>(0.72-1.04)</td>
<td>0.121</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>65 223 214</td>
<td>0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

SNP: Single nucleotide polymorphism, OR: Odds ratio, 95% CI: 95% confidence interval, HWE: Hardy-Weinberg equilibrium. P* value from were calculated from two-sided Chi-squared test/Fisher's exact test.
ZFHX3 polymorphism and stroke

Table 3. Frequency distributions of prominent SNPs and their associations with the risk of developing stroke

<table>
<thead>
<tr>
<th>SNP</th>
<th>Minor allele</th>
<th>MAF</th>
<th>Case OR (95% CI)</th>
<th>Control OR (95% CI)</th>
<th>p*</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs16971384</td>
<td>G</td>
<td>367</td>
<td>0.92 (0.71-1.20)</td>
<td>0.544</td>
<td>0.544</td>
</tr>
<tr>
<td>rs17680796</td>
<td>T</td>
<td>270</td>
<td>0.87 (0.68-1.12)</td>
<td>0.288</td>
<td>0.121</td>
</tr>
<tr>
<td>rs13336412</td>
<td>A</td>
<td>411</td>
<td>0.73 (0.56-0.95)</td>
<td>0.021*</td>
<td>0.021*</td>
</tr>
<tr>
<td>rs9921395</td>
<td>A</td>
<td>85</td>
<td>0.72 (0.30-0.99)</td>
<td>0.046</td>
<td>0.046*</td>
</tr>
<tr>
<td>rs17681554</td>
<td>C</td>
<td>175</td>
<td>0.85 (0.66-1.11)</td>
<td>0.230</td>
<td>0.230</td>
</tr>
<tr>
<td>rs879324</td>
<td>A</td>
<td>311</td>
<td>0.88 (0.69-1.13)</td>
<td>0.331</td>
<td>0.331</td>
</tr>
</tbody>
</table>

*P ≤ 0.05 indicates statistical significance. P* values were calculated by unconditional logistic regression.

We want to figure out the sequence variants if it increased the risk of stroke in Chinese Han descent. Previously, a genome-wide association study (GWAS) in patients of European population identified that risk variants in ZFHX3 on 16q22 were related with stroke and atrial fibrillation [19]. Thus Six potentially functional SNPs in ZFHX3 were selected for genotyping. In this case-control study in northwest Chinese Han population to verify positive association signals identified in the previous GWAS findings, ZFHX3 has been reported to confer increased risk of AF in European and Chinese populations [20]. All SNPs located in the intron region of zinc finger homeobox 3 (ZFHX3) gene on chromosome 16q22, also known as AT motif-binding factor 1 (ATBF1) or AT-binding transcription factor (ATBT). ZFHX3 is expressed in various tissues including heart, liver, lung, kidney, pituitary gland and brain. The gene expression could regulate growth and differentiation in several tissues, including differentiation in neuronal and skeletal muscle [21]. In our study, none of genotype association was identified. Notably, we assessed the association of ZFHX3 with stroke and finally identified SNPs rs13336412 and rs9921395 in the ZFHX3 gene in Dominant model associated with stroke with an OR 0.7301 (P = 0.02) and 0.7234 (P = 0.046), indicating that rs13336412 or rs9921395 may play an critical role as a protected factor in pathogenesis of stroke.

ZFHX3 encodes a transcription factor with multiple homeodomains and zinc finger motifs, and multiple transcript variants expressed from alternate promoters and encoding different isoforms have been found for this gene. The ZFHX3 gene was identified in AF and subsequently associated with Ischemic Stroke and cardioembolism subtypes [20]. The PITX2 gene closest to the AF and IS risk variants (tagged by rs2200733 and rs10033464) as a risk factor which associated strongly with cardioembolic stroke is also observed in noncardiogenic stroke [8]. The susceptibility of stroke, especially the ischemic stroke, increased by AF, which is an important risk factor for morbidity and mortality in China and the rest of world [22-24]. Therefore, we hypothesis ZFH- X3 is not only associated with stroke resulting from AF or Ischemic stroke but also associated with other kind of stroke subtypes in general. Little is known about the association of ZFHX3 and stroke in the previous population-based study, but it is verified ZFHX3 had strong association with AF while AF is one of the most stroke risk factors that confers a 4 to 5-fold increasing risk among patients with stroke [6]. There is available effective interventions to lower the risk, disappointingly, the patients who has the highest risk for a new stroke still didn’t attract enough attention for the ignorance of subtle relevancy between AF and stroke in clinical health care, not only in developed country like Sweden [3], but in the Asian-Pacific region especially in China.

Even though our study fail to reveal significant association between stroke and ZFHX3, Bellenguez et al found a significant association in the ZFHX3 gene between the rs7193-343 polymorphisms and Ischemic Stroke of the cardioembolism subtype in European descent. Then Wang et al replicated that the rs7193343 polymorphism in the ZFHX3 gene had significant linkage to Ischemic Stroke in the Northern Chinese Han Population [25]. Traway et al verified previous associations for cardioembolic stroke near ZFHX3 (P = 2.28×10^−6).
ZFHX3 polymorphism and stroke

Bellenguez et al also confirmed an association between cardioembolic stroke and a SNP in the ZFHX3 gene, which was initially associated with atrial fibrillation, a well-recognized risk factor for stroke [27]. Therefore, the role of rs7193834 may vary among different populations, with the ZFHX3 gene suggested to confer a significant risk of AF in the Chinese Han population. Thus, further research is needed to assess the relationship between ZFHX3 and different stroke subtypes among populations lived in other region of China.

To the best of our knowledge, it is the first study on association between the ZFHX3 and stroke risk in northwest Chinese Han population yet previous findings are focused on the variants AF on ZFHX3 which might ignore the effect of susceptibility loci in Chinese stroke patients. Though it is hard to judge ZFHX3 associated with stroke performed by our research, the subtle association between stroke and ZFHX3 still need to be revealed in the following study for AF conferring 4 to 5-fold increasing risk among stroke patients.

Our studies had some limitations. Firstly, the analyses testing for subtypes of stroke for the limitation of sample size and we unable to collect detailed diagnostic data from the control group, because some of them enrolled in the study had not obtain medical history like hypertension, CAD, stroke or diabetes mellitus in all controls, so correction for this variant was impossible, this may leads to selection bias for the control group. Secondly, the ethnicity of study participants was limited to the northwest Chinese Han population but majority of them lived in Shanxi province and nearby area. Therefore further meta-analysis is required in other research to confirm validation and applicability to other ethnic groups within Chinese Han population. Our findings may also not be applicable to individuals of other races and ethnicities. Finally, the function genetic variants and mechanisms underpinning this association will need well-designed studies in additional large-scale Chinese Han populations including fine mapping and laboratory studies.

To sum up, our study investigated the role of genetic variants of ZFHX3 in stroke and the significant association between rs16971384, rs12596992, rs17680796, rs17681554 and rs879324 in ZFHX3. However, further large and functional studies are needed to confirm our findings.

Acknowledgements

This work is supported by the National 863 High-Technology Research and Development Program (No. 2012AA02A519).

Disclosure of conflict of interest

None.

Address correspondence to: Tianbo Jin, National Engineering Research Center for Miniaturized Detection Systems, Mailbox: 386, 229 North Taibai Road, Xi’an 710069, China. Tel: +86-29-88303800; Fax: +86-29-88303800; E-mail: jintianbo@gmail.com

References

ZFHX3 polymorphism and stroke


[15] Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ and Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007; 81: 559-575.


ZFHX3 polymorphism and stroke

