Gene polymorphisms in association with self-reported stroke in US adults

Amy Z Fan¹
Jing Fang¹
Ajay Yesupriya²
Man-huei Chang²
Greta Kilmer¹
Meaghan House³
Donald Hayes¹
Renée M Ned²
Nicole F Dowling²
Ali H Mokdad¹

¹National Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control and Prevention, Atlanta, GA, USA; ²Office of Public Health Genomics, Coordinating Center for Health Promotion, Centers for Disease Control and Prevention, Atlanta, GA, USA; ³School of Public Health, Emory University, Atlanta, GA, USA

Purpose: Epidemiologic studies suggest that several gene variants increase the risk of stroke, and population-based studies help provide further evidence. We identified polymorphisms associated with the prevalence of self-reported stroke in US populations using a representative sample.

Methods: Our sample comprised US adults in the Third National Health and Nutrition Examination (NHANES III) DNA bank. We examined nine candidate gene variants within ACE, F2, F5, ITGA2, MTHFR, and NOS3 for associations with self-reported stroke. We used multivariate regression and Cox proportional hazards models to test the association between these variants and history of stroke.

Results: In regression models, the rs4646994 variant of ACE (I/I and I/D genotypes) was associated with higher prevalence adjusted prevalence odds ratio [APOR] = 2.66 [1.28, 5.55] and 2.23 [1.30, 3.85], respectively) compared with the D/D genotype. The heterozygous genotype of MTHFR rs1801131 (A/C) was associated with lower prevalence of stroke (APOR = 0.48 [0.25, 0.92]) compared with A/A and C/C genotypes. For rs2070744 of NOS3, both the C/T genotype (APOR = 1.91 [1.12, 3.27]) and C/C genotype (APOR = 3.31 [1.66, 6.60]) were associated with higher prevalence of stroke compared with the T/T genotype.

Conclusion: Our findings suggest an association between the prevalence of self-reported stroke and polymorphisms in ACE, MTHFR, and NOS3 in a population-based sample.

Keywords: stroke, gene, polymorphisms, NHANES III, gene association analysis

Introduction

Stroke is the third leading cause of mortality in the United States and can cause serious long-term disability among survivors. Stroke can be classified into two major types: ischemic (87% of all strokes in the United States) and hemorrhagic, and is associated with atherosclerosis of large blood vessels or occlusion of small arteries in the brain. Previous epidemiologic studies indicate that stroke is a multifactorial disease caused by a complex interaction between genes and the environment. Ischemic and hemorrhagic stroke have several risk factors in common, including hypertension, hyperlipidemia, diabetes, and smoking. However, a substantial portion of patients that suffer from stroke do not have these known risk factors.

Studies of twins, siblings, and families indicate that genetic factors might contribute to the risk of stroke. Numerous studies have examined the association between individual genes and the increased risk of stroke, primarily focusing on genes involved in thrombosis or coagulation. A meta-analysis by Casas et al found four gene variants that were consistently associated with ischemic stroke, including angiotensin-converting enzyme (ACE) insertion/deletion, prothrombin (F2) G20210A, factor V (F5) Leiden A506Q,
and methylenetetrahydrofolate reductase (MTHFR) C677T. Based on a detailed examination of the current literature, Bersano et al. also found that these were among the most likely candidate genes related to stroke, and suggest that extensive investigations are needed to confirm these associations.

Although several previous studies found no association between integrin alpha 2 (ITGA2) gene variants and stroke, a recent study of North Americans demonstrated that this gene is in fact associated with an increased risk for ischemic stroke. Research on the association between nitric oxide synthase 3 (NOS3) and stroke is also inconclusive; however, variants within this gene have been shown to confer risk among African Americans.

Identifying common gene variants that might increase the risk of stroke is of public health importance because it could allow identification and targeting of subpopulations that might be at an increased risk for the disease. Although previous epidemiologic studies have suggested that several common gene variants contribute to the increased risk of stroke, large-scale population-based studies could provide stronger evidence on those associations. The purpose of the present study is to assess if polymorphisms within ACE, F2, F5, ITGA2, MTHFR, and NOS3 are associated with the prevalence of self-reported stroke in a large, nationally representative sample of US adults.

**Materials and methods**

**Study sample**

The Third National Health and Nutrition Examination Survey (NHANES III) is a complex, multistage sample survey conducted by the National Center for Health Statistics of the Centers for Disease Control and Prevention from 1988 through 1994. This cross-sectional study was designed to provide national estimates of common diseases and their respective risk factors among the civilian non-institutionalized population aged two months or older in the United States. Data collection for NHANES occurred at three levels: a brief household screener interview, an in-depth household survey interview, and a medical examination. Population weights were calculated for each individual to make the data representative of the US population. In the second phase of NHANES III, from 1991 through 1994, white blood cells from participants aged 12 years or older were frozen and cell lines were immortalized using the Epstein-Barr virus, creating a DNA bank. The analysis was performed among adults aged 17 years and older (n = 5973) using the data collected in this DNA bank. This study was approved by the National Center for Health Statistics Ethics Review Board.

**Candidate genes and genotyping methods**

The candidate gene variants selected for current analysis (nine variants in six genes) were ACE (rs4646994), F2 (rs1799963), F5 (rs6025), ITGA2 (rs1126643), MTHFR (rs1801133, rs1801131, rs2066470), and NOS3 (rs1799983, rs2070744). Genotypes were assayed either by TaqMan (5’ nuclease assay; Applied Biosystems, Foster City, CA) or by the MGB Eclipse Assay (3’ hybridization-triggered fluorescence reaction; Nanogen, Bothell, WA). Water controls and DNA samples with known genotypes (from Coriell Cell Repository, Camden, NJ) were included in each well. Deviations of Hardy-Weinberg proportions were tested using unweighted chi-square goodness-of-fit tests. Complete descriptions of the genotyping and quality control methods have been previously published.

**Cases – self-reported history of stroke**

Participants were classified as having experienced a stroke if they answered “yes” to the question “Has a doctor ever told you that you had a stroke?” They were then asked “How old were you when you were first told you had a stroke?”

**Statistical analysis**

All analyses accounted for the NHANES III sampling design using specialized procedures available in SAS-callable SUDAAN 9.01 (Research Triangle Institute, Research Triangle Park, North Carolina) for the analysis of complex surveys. All models included sample weights that were recalculated for the NHANES III DNA bank data. We used the Taylor series linearization approach, implemented in SUDAAN, to calculate standard errors that account for correlations in the data due to the sampling design, including the possible genetic relatedness of individuals sampled from the same household.

We used the Satterthwaite-adjusted F-statistics (available in the RLOGIST procedure) to test the association of stroke with the nine selected gene variants. Multivariate regression models were used to examine the association between self-reported stroke prevalence and study gene variants, adjusting for potential confounders including age, sex, race/ethnicity, and education. Other previously identified stroke risk factors were either 1) not associated with the genes in this study or 2) involved in the biological pathway between gene and disease. Therefore, no other stroke risk factors were added to the models. A codominant model of inheritance was assumed for all variants except for rs2066470 (MTHFR), rs1799963 (F2), and rs6025 (F5), which were tested under a dominant model due to their low minor allele frequencies in the population (less than 5%). Adjusted prevalence odds
ratios (APORs) and 95% confidence intervals were estimated from these models. Additionally, we used Cox proportional hazards models, which have been shown to increase statistical power in cross-sectional genetic association studies, with self-reported age at stroke as the outcome. The results from Cox models and logistic regression models are compared.

Results

Basic characteristics of the study sample are presented in Table 1. Our sample comprised 5973 US adults aged 17 years or older, of whom 156 reported having had a stroke. Participants reporting stroke were older and more likely to have a lower level of education than controls.

The distribution of genes variants between stroke cases and controls (Table 2) showed that for rs4646994 (ACE), participants reporting stroke were more likely to have the I/I and I/D genotypes than controls (P = 0.052). For rs1801131 (MTHFR), controls were more likely to have the A/C genotype than participants with a history of stroke (P = 0.036). Finally, participants with a history of stroke were more likely to have rs2070744 (NOS3) C/C or C/T genotypes than controls (P = 0.020). No difference in the distribution of other gene variants was observed between stroke cases and controls.

In multivariate regression models, gene variants of ACE, MTHFR, and NOS3 were significantly associated with stroke after controlling for age, sex, race/ethnicity, and education (Table 3). In a codominant model, the rs4646994 variant of ACE (I/I and I/D genotypes) was significantly associated with higher prevalence odds of stroke adjusted prevalence odds ratio [APOR] = 2.66 [1.28, 5.55] and 2.23 [1.30, 3.85], respectively) as compared to the D/D genotype. In a codominant model, the heterozygous MTHFR genotype of rs1801131 (A/C) was significantly associated with a lower prevalence of stroke (APOR = 0.48 [0.25, 0.92]) compared with the A/A and C/C genotypes. Additionally, the variant rs2070744 of NOS3 was significantly associated with stroke.

Table 1 Basic characteristics of individuals aged 17 years or older in the NHANES III DNA bank

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Cases</th>
<th>Controls</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean or percent (SE)</td>
<td>N = 156</td>
<td>N = 5817</td>
<td></td>
</tr>
<tr>
<td>Mean age (yrs)</td>
<td>67.7 (1.57)</td>
<td>43.5 (0.68)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Sex (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>55.9 (4.41)</td>
<td>52.0 (0.77)</td>
<td>0.3900</td>
</tr>
<tr>
<td>Men</td>
<td>44.1 (4.41)</td>
<td>48.0 (0.77)</td>
<td></td>
</tr>
<tr>
<td>Race/ethnicity (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White, non-Hispanic</td>
<td>82.3 (2.96)</td>
<td>81.6 (1.73)</td>
<td>0.0600</td>
</tr>
<tr>
<td>Black, non-Hispanic</td>
<td>14.8 (2.69)</td>
<td>12.4 (1.58)</td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>2.9 (0.99)</td>
<td>6.0 (0.86)</td>
<td></td>
</tr>
<tr>
<td>Education (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-high school graduate</td>
<td>41.7 (5.93)</td>
<td>22.3 (1.40)</td>
<td>0.0002</td>
</tr>
<tr>
<td>High school graduate</td>
<td>34.8 (5.28)</td>
<td>34.6 (1.28)</td>
<td></td>
</tr>
<tr>
<td>Some college</td>
<td>15.2 (4.25)</td>
<td>21.4 (1.51)</td>
<td></td>
</tr>
<tr>
<td>Bachelor degree or higher</td>
<td>8.3 (2.61)</td>
<td>21.6 (1.51)</td>
<td></td>
</tr>
</tbody>
</table>

*P values were obtained from chi-square tests for categorical variables and t-tests for continuous variables.
Both the C/T genotype (APOR = 1.91 [1.12, 3.27]) and C/C genotype (APOR = 3.31 [1.66, 6.60]) were associated with higher prevalence odds of stroke in comparison with the T/T genotype. The remaining gene variants tested were not found to be significantly associated with stroke. The results from logistic regression models were consistent with those from Cox proportional hazards models.

### Discussion

We used a large representative sample of the US population to investigate 9 variants in 6 genes that were important factors in the pathophysiology of stroke and were reported in previous studies to be associated with stroke. In this study we found that 1) the rs4646994 variant of ACE (I/I and I/D genotypes) was significantly associated with a higher prevalence of stroke as compared with the D/D genotype; 2) the heterozygous genotype (A/C) of MTHFR variant rs1801131 was significantly associated with a lower prevalence of stroke as compared with the A/A and C/C genotypes; and 3) both the C/T genotype and the C/C genotype of the variant rs2070744 of NOS3 were associated with a higher prevalence of stroke in comparison with the T/T genotype.

Angiotensin-converting enzyme (ACE) converts angiotensin I to angiotensin II, which is a vasoconstrictor that contributes to vascular hypertrophy and the development of atherosclerosis. Plasma and intracellular levels of ACE are partially determined by the presence of the ACE I/D polymorphism (rs4646994). The D/D genotype of the ACE I/D polymorphism has been found in multiple studies to be associated with increased risk for ischemic stroke in both European and Asian populations. However, our results are in the opposite direction: study participants that were homozygous for the D allele were at decreased risk for stroke.

Methylenetetrahydrofolate reductase (MTHFR) is a key metabolic enzyme that is involved in the conversion of homocysteine to methionine. The common C677T polymorphism of MTHFR (rs1801133) is frequently studied in association with arteriosclerosis because the homozygous variant is associated with reduced activity of the enzyme, leading to increased levels of homocysteine. Increased levels of homocysteine might be involved in the pathogenesis of atherosclerosis and a risk factor for stroke, as patients with markedly elevated homocysteine levels have been shown to develop premature vascular disease. In a meta-analysis, Casas et al showed a statistically significant association between ischemic stroke and the C677T polymorphism. However, the study did not include the MTHFR gene polymorphism A1298C (rs1801131), which we found to be associated with decreased risk of stroke in the A/C heterozygotes. The A1298C polymorphism has been studied to a lesser extent, and an interaction between the two polymorphisms has been proposed. The biochemical profile of A1298C/C677T compound heterozygotes has been shown to be similar to 677T homozygotes.
The endothelial nitric oxide synthase (NOS3) catalyses nitric oxide (NO) synthesis, which relaxes smooth muscle cells, reduces smooth muscle growth, and inhibits platelet and leukocyte adhesion to the vascular endothelium. Variants in NOS3 have been studied in association with increased risk of cardiovascular disease.\(^2\)-\(^4\) In particular, the polymorphism NOS3 T-786C (rs2070744) is associated with increased susceptibility to coronary vasospasm in individuals expressing the variant C allele.\(^5\) Khurana et al suggested a biological mechanism for aneurysm physiology and found that the T-786C genotype influenced the size at which an aneurysm ruptures.\(^6\) Although polymorphisms in NOS3 have been associated with an increased risk of ischemic heart disease, there are inconsistent results for the risk of stroke.\(^7\) Most studies showing an association with stroke involved African American participants. Studies yielding inconsistent results might have had an insufficient number of participants of African descent to detect an association.\(^8\)-\(^9\) To confirm the association, this type of study will need to be replicated in a larger sample of African Americans.

**Limitations**

Our study has several limitations. A main concern with association studies is that a priori knowledge of the putative risk polymorphisms to be studied is required. To address this, we have cited previous meta-analyses which show substantial evidence that these variants are associated with stroke. Also, when a genetically heterogeneous population is studied, racial/ethnic variations in allele frequencies create the possibility of spurious associations.\(^10\) To address this concern, we used both logistic regression models and Cox proportional hazards models. Hazards models give more conservative effect estimates than logistic regression models when the incidence of disease is high. We were able to replicate significant associations using hazard ratios, which supports the results of the logistic regression analyses in our study.\(^11\)

The limitations of NHANES III include self-reported outcome (stroke) without clinical confirmation and with no discrimination between ischemic and hemorrhagic subtypes of stroke. Some researchers argue that these different pathological conditions are unlikely to be under the same genetic influence.\(^12\) Additionally, persons who died from stroke or were unable to take part in the study because of disability were not included in the sample. For this reason, we expect that the cases of stroke in this study are less severe than those in the general population.

**Future studies**

While our study sought to identify polymorphisms that were associated with the prevalence of self-reported stroke in a large-scale population-based association study, these results should be interpreted with caution. Demonstration of an association is not by itself sufficient evidence for a causative role of the gene variant studied.\(^13\) Furthermore, stroke is a highly complex disease, and our study was not able to examine the different subtypes of stroke in relation to gene polymorphisms. Intermediate markers of stroke (eg, atherosclerosis or intima-media thickness) or risk factors (hypertension, hyperlipidemia, and diabetes) might also be under tight genetic control and could be a more precise means of detecting an association rather than the related disease outcome. Future population-based studies should assess the polymorphisms that are specific to each stroke subtype, as well as the associations in intermediate markers of stroke. Further studies should also examine gene-gene and gene-environment interactions among population subgroups for complex conditions such as stroke.\(^14\)

**Conclusion**

In conclusion, our study found associations between stroke and gene variants within ACE, MTHFR, and NOS3 using a nationally representative, population-based survey. Identifying such variants might allow for a better understanding of the pathogenesis of stroke and possibly the improved targeting and treatment of individuals or populations at an increased risk for stroke. A genome-wide association study using the cohort from the same population with clear clinical diagnosis of stroke can be used to verify or confirm the findings.

**Disclosures**

The findings and conclusions in this article are those of the authors and do not necessarily represent the views of the Centers for Disease Control and Prevention.

The authors declare no conflicts of interest.

**References**


