Genome-Wide Association Study of Intracranial Aneurysms Confirms Role of Anril and SOX17 in Disease Risk

Tatiana Foroud, PhD; Daniel L. Koller, PhD; Dongbing Lai, MS; Laura Sauerbeck, RN, MS; Craig Anderson, MD; Nerissa Ko, MD; Ranjan Deka, PhD; Thomas H. Mosley, PhD; Myriam Fornage, PhD; Daniel Woo, MD, MS; Charles J. Moomaw, PhD; Richard Hornung, DPH; John Huston, MD; Irene Meissner, MD; Joan E. Bailey-Wilson, PhD; Carl Langefeld, PhD; Guy Rouleau, MD, PhD; E. Sander Connolly, MD; Bradford B. Worrall, MD, MSc; Dawn Kleindorfer, MD; Matthew L. Flaherty, MD; Sharyl Martini, MD, PhD; Jason Mackey, MD, MS; Felipe De Los Rios La Rosa, MD; Robert D. Brown, Jr, MD; Joseph P. Broderick, MD; the FIA Study Investigators

Background—Genome-wide association studies have identified novel genetic factors that contribute to intracranial aneurysm (IA) susceptibility. We sought to confirm previously reported loci, to identify novel risk factors, and to evaluate the contribution of these factors to familial and sporadic IA.

Method—We utilized 2 complementary samples, one recruited on the basis of a dense family history of IA (discovery sample 1: 388 IA cases and 397 controls) and the other without regard to family history (discovery sample 2: 1095 IA cases and 1286 controls). Imputation was used to generate a common set of single nucleotide polymorphisms (SNP) across samples, and a logistic regression model was used to test for association in each sample. Results from each sample were then combined in a meta-analysis.

Results—There was only modest overlap in the association results obtained in the 2 samples. In neither sample did results reach genome-wide significance. However, the meta-analysis yielded genome-wide significance for SNP on chromosome 9p (CDKN2BAS, rs6475606, \( P = 3.6 \times 10^{-8} \)) and provided further evidence to support the previously reported association of IA with SNP in SOX17 on chromosome 8q (rs1072737, \( P = 8.7 \times 10^{-5} \)). Analyses suggest that the effect of smoking acts multiplicatively with the SNP genotype, and smoking has a greater effect on risk than SNP genotype.

Conclusion—In addition to replicating several previously reported loci, we provide further evidence that the association on chromosome 9p is attributable to variants in CDKN2BAS (also known as ANRIL, an antisense noncoding RNA). (Stroke. 2012;43:00-00.)

Key Words: genome-wide association study ■ intracranial aneurysm

The risk of intracranial aneurysm (IA) is increased among individuals with first-degree relatives with history of IA. Many approaches have been used to identify genes contributing to the risk of IA. Genome-wide association studies identified and replicated associations on chromosome 4q31.23 (EDNRA), 8q12.1 (SOX17), 9p213 (CDKN2A/CDKN2B/CDKN2BAS), 10q24.32 (CNNM2), 12q22, 13q13.1 (KL/STARD13), 18q11.2 (RBP5), and 20p12.1.3-6 Most studies utilized sporadic IA cases; familial cases may potentially involve different genetic risk factors. The present study was undertaken to compare and contrast results from a genome-wide association study of familial and sporadic IA cases.

Subjects and Methods

Discovery Sample 1
Discovery sample (DS) 1 consisted of individuals recruited through the Familial Intracranial Aneurysm (FIA) I study, which recruited familial cases appropriate for linkage analysis through an international consortium and applied rigorous inclusion and exclusion criteria. A set of independent unrelated cases was obtained by

Received March 5, 2012; accepted August 7, 2012.
From the Indiana University School of Medicine, Indianapolis, IN (T.F., D.L.K., D.L., J.M.); University of Cincinnati College of Medicine, Cincinnati, OH (L.S., R.D.; D.W., C.J.M., R.H., D.K., M.L.F., S.M., F.D.L.R.L.R., J.B.P.); the George Institute for Global Health of Royal Prince Alfred Hospital and University of Sydney, Sydney, Australia (C.A.); University of California, San Francisco, San Francisco, CA (N.K.); University of Mississippi Medical Center, Jackson, MS (T.H.M.); University of Texas Health Science Center, Houston, TX (M.F.); Mayo Clinic, Rochester, MN (J.H., I.M., R.D.B.); National Human Genome Research Institute, Baltimore, MD (J.E.B.W.); Wake Forest University, Winston-Salem, NC (C.L.); University of Montreal, Montréal, Québec, Canada (G.R.); Columbia University School of Medicine, New York, NY (E.S.C.); University of Virginia School of Medicine, Charlottesville, VA (B.B.W.).

The online-only Data Supplement is available with this article at http://stroke.ahajournals.orglookup/suppl/doi:10.1161/STROKEAHA.112.656397/DCl.

Correspondence to Tatiana Foroud, PhD, Indiana University School of Medicine, 410 W. 10th Street, Indianapolis, IN 46202. E-mail tforoud@iupui.edu

© 2012 American Heart Association, Inc.

Stroke is available at http://stroke.ahajournals.org

DOI: 10.1161/STROKEAHA.112.656397
selecting 1 self-reported white individual with definite IA from each family (n=389). White control subjects were obtained from 2 population-based studies, Genetic and Environmental Risk Factors for Hemorrhage Stroke and the Cincinnati Control Cohort. All studies were approved by the appropriate Institutional Review Boards. Additional information is provided in the Online Supplement.

**DS2**

During FIA II study recruitment, the requirement for family history of IA was removed and both familial and sporadic IA cases were enrolled. The same exclusion criteria were in place, and all case subjects underwent the same rigorous review. 829 white IA case subjects from FIA II and 61 white sporadic aneurysmal subarachnoid hemorrhage case subjects were selected.

The sample was augmented by white case subjects and control subjects identified from other studies. Subjects with incident cases of subarachnoid hemorrhage secondary to documented or presumed ruptured IA (n=160) who were frequency-matched to control subjects (n=128) were obtained from the Australasian Cooperative Research on Subarachnoid Hemorrhage Study (ACROSS). An additional 184 white IA case subjects were selected from a prospective San Francisco cohort study of adult patients with spontaneous subarachnoid hemorrhage attributable to IA confirmed by noncontrast computed tomography and cerebral angiogram. All studies were approved by the appropriate Institutional Review Boards.

Genotypic data from 1148 white control subjects were obtained through a collaborative agreement with the Atherosclerosis Risk in Communities (ARIC) study. In the ARIC sample, a subset of subjects who never had a stroke or transient ischemic attack was matched to the DS2 case subjects by sex and age (±5 years).

**Genotyping, Quality Review, and Imputation**

Genotyping was performed using the Axiom array for all samples except the ARIC control subjects. All released genotypes underwent a common quality review pipeline. A principal component analysis was performed to identify and remove samples from subjects with African, Asian, or Hispanic admixture (n=47). Genotypic data for the ARIC samples were obtained from the Affy 6.0 array and underwent a similar quality review.

Imputation for the Axiom samples was performed for all autosomes using IMPUTE2. All samples genotyped on the Axiom array (n=2115) were imputed together using the 1000 Genomes haplotypes (n=1094) as the phased reference panel. The ARIC samples were included as another unphased reference panel to maximize available information at each imputed single nucleotide polymorphism (SNP). We did not impute genome-wide SNP genotypes for the ARIC sample. Extensive and detailed quality review was performed to ensure that spurious association was not detected based on platform effects. Using an aggressive filtering approach, we retained 453 699 SNP for analysis of DS2.

**Statistical Analysis**

The genotyped and imputed SNP were used to test for association with IA susceptibility using a logistic regression model. No additional covariates were necessary. Data from DS1 were analyzed initially using only the genotyped SNP. Data from DS2 were analyzed using the genotyped data from ARIC, along with both genotyped and imputed data from the samples genotyped on the Axiom array using the SNPTEST version 2 software. Imputed genotypes were encoded in the logistic model as the expected allele count.

Meta-analysis was performed by combining results from DS1 and DS2 for the common set of 451 088 SNP genotyped or imputed in both samples. METAL was used using an inverse-variance weighting scheme. Genomic control was used before and after meta-analysis to down-weight results in each sample when the genomic inflation coefficient exceeded 1.00. We evaluated results to identify findings meeting genome-wide significance criteria ($P<5\times10^{-8}$), and we also sought to replicate previous associations using a more modest significance level ($P<10^{-4}$).

To test whether there might be >1 risk variant in a particular gene or gene region contributing to the association, we performed conditional analyses. For the statistically significant region on chromosome 9, we identified the SNP with the most extreme probability value in the meta-analysis. We then modified the logistic regression model to include the genotype at the most significant SNP in the region in each of the samples and performed meta-analysis to combine the results from each sample. Finally, we reviewed the probability value for the other SNP in the region to determine whether any other SNP remained statistically significant.

<table>
<thead>
<tr>
<th>Table 1. Sample Demographics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Discovery Sample 1</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Cases</strong></td>
</tr>
<tr>
<td>Case number</td>
</tr>
<tr>
<td><strong>Discovery Sample 2</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Cases</strong></td>
</tr>
<tr>
<td>Case number</td>
</tr>
</tbody>
</table>

Table 1. Sample Demographics
Because cigarette smoking is a strong risk factor in IA,\textsuperscript{16,17} we examined the possible interaction of the most highly associated SNP on chromosomes 8 and 9 with cigarette smoking. A logistic regression model, with age and sex as covariates, was used to test for departures from a multiplicative relationship between the 3 genotypes for that SNP and cumulative exposure to smoking, as measured by pack-years. Because the distribution of pack-years was highly skewed, we used the log of pack-years in the logistic model, with 0.05 assigned to never-smokers. Results from DS1 and DS2 were combined in a meta-analysis weighted by the inverse variance of the sample estimates.

Figure 1. Results of the genome-wide association analysis for intracranial aneurysm (IA) susceptibility. A, Discovery sample 1. B, Discovery sample 2. C, Meta-analysis.
Table 2. Summary of Top Association Results

<table>
<thead>
<tr>
<th>Chr</th>
<th>Position (bp)</th>
<th>SNP</th>
<th>Gene (Nearest Gene)</th>
<th>EA</th>
<th>AF</th>
<th>Axiom (G/I)</th>
<th>Discovery Sample 1</th>
<th>Discovery Sample 2</th>
<th>Meta-Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>P Value</td>
<td>OR</td>
<td>P Value</td>
</tr>
<tr>
<td>2</td>
<td>128796641</td>
<td>rs11693075</td>
<td>SAP130 (11kb)</td>
<td>A</td>
<td>0.75</td>
<td>G</td>
<td>0.8008</td>
<td>1.03</td>
<td>9.85×10⁻⁶</td>
</tr>
<tr>
<td>2</td>
<td>183026612</td>
<td>rs1897472</td>
<td>PDE1A</td>
<td>A</td>
<td>0.69</td>
<td>G</td>
<td>6.66×10⁻⁷</td>
<td>1.78</td>
<td>0.24*</td>
</tr>
<tr>
<td>2</td>
<td>183043380</td>
<td>rs3769801</td>
<td>PDE1A</td>
<td>G</td>
<td>0.69</td>
<td>G</td>
<td>5.53×10⁻⁷</td>
<td>1.79</td>
<td>0.13*</td>
</tr>
</tbody>
</table>

Meta-Analysis

Combining the results from DS1 and DS2 in a meta-analysis yielded genome-wide significance for the region on chromosome 9p (CDKN2BAS; rs6475606; P=3.6×10⁻⁸; Figure 1C). No other region attained this level of significance. We systematically compared our results with those in previous genome-wide association studies and replicated not only the association on chromosome 9p but also the association with SOX17 on chromosome 8q (rs1072737; P=8.7×10⁻⁵; Figure 2). We did not find substantial support for the previously reported associations in other chromosomal regions (Supplemental Figure I). However, at least on chromosome 20, our study did not include any SNP with at least moderate linkage disequilibrium with the most significant SNP previously reported. Therefore, we cannot definitely evaluate whether our sample provides support for the previously reported association.

Conditional Analysis

Given previous reports of association of several genes in the chromosome 9 region, we performed conditional analysis to test for evidence of >1 risk factor in the region. We performed analyses in DS1 and DS2 separately, including the most significant SNP (rs6475606) in the logistic regression model. After meta-analysis of the results in the 2 samples in this region, no SNP remained significant when rs6475606 was included in the model (all P>0.54), suggesting there is only 1 risk factor for IA in this chromosomal region.

Discussion

We found the strongest evidence of association with a SNP in CDKN2BAS, also known as ANRIL. Previous studies reported an association of IA,²⁻⁶,¹⁸,¹⁹ as well as myocardial infarction,¹⁸ large-vessel ischemic stroke subtype,²⁰ and aortic aneurysm,¹⁸,²¹

Regions

Results

DS1

The final sample included 388 familial IA case subjects and 397 control subjects (Table 1) with genotypes for 492,284 SNP. Results are summarized in Figure 1A and Table 2. Although no results achieved traditional criteria for genome-wide significance, several SNP yielded interesting results within phosphodiesterase 1A (PDE1A; rs3769801; P=5.5×10⁻⁷), with little evidence of increased false-positive results (λ=1.047).

DS2

The final sample included 1095 IA case subjects and 1286 control subjects (Table 1) with genotypes for 453,699 SNP (Figure 1B and Table 2). Although no results attained genome-wide significance, 2 regions had SNP with genome-wide significance, 2 regions had SNP with moderate linkage disequilibrium with the most significant SNP previously reported. Therefore, we cannot definitely evaluate whether our sample provides support for the previously reported association.

Gene and Smoking Relationship

Statistical testing of the SNP on chromosomes 8 (rs1072737) and 9 (rs6475606) did not find evidence for an interaction of either SNP with log (pack-years). Rather, the data were consistent with a multiplicative relationship between the SNP genotype and smoking (Table 3). Importantly, when both genotype and smoking are modeled together, the effect of smoking on disease risk remains substantially greater than that of the SNP genotypes (smoking odds ratio, 2.5; SNP odds ratio, 1.25–1.36).
with SNP in this region. Others examined the association in multiplex families as well as sporadic IA and found consistent evidence of association with rs1333040 (allele T), located in intron 12 of CDKN2BAS and having limited linkage disequilibrium spanning introns 7 through 15 of CDKN2BAS. The association of this SNP with both sporadic and familial IA parallels our findings. In DS1, we found evidence for association with the high-risk T allele (\(P = 0.02\); odds ratio, 1.26). We imputed this SNP in the ARIC sample and saw even greater evidence of this association in DS2 (\(P = 1.4 \times 10^{-5}\); odds ratio, 1.29). When the 2 samples were combined, the association strengthened (\(P = 7.6 \times 10^{-6}\); odds ratio, 1.29).

The linkage disequilibrium structure of the chromosome 9p21 region has at least 2 major linkage disequilibrium blocks. The first is associated with vascular diseases and the other is associated with diabetes mellitus. SNP in the region are associated with several types of cancer as well as open-angle glaucoma.\(^{22}\) Using conditional analyses in our sample, we did not find evidence of \(>1\) risk factor for IA within this chromosomal region.

ANRIL is a long intergenic noncoding RNA. A mutant mouse with a deletion corresponding to this region has significantly greater suppression of the RNA encoded by Cdkn2a and Cdkn2b.\(^{23}\) Cultured aortic smooth muscle cells

Figure 2. Comparison with selected previously reported results from genome-wide association studies.\(^{5,6}\) A, Chromosome 8q12.1. B, Chromosome 9p213. When the most significant single nucleotide polymorphism (SNP) from the initial report was available in the meta-analysis, a white diamond indicates the SNP, and the SNP is listed at the top of the Figure. If the SNP was not available in the meta-analysis sample, then an arrow is used to denote the position of that SNP. Each circle indicates the probability value for a SNP at that position in the meta-analysis. The color of the square symbol denotes the extent of linkage disequilibrium (as computed by \(r^2\)) with the SNP reported in the initial report.
from these mutant mice had increased proliferative activity. These data suggest that the association of ANRIL, with both atherosclerosis and IA may be attributable to similar pathophysiologic mechanisms, because both involve thinning of the tunica media layer in affected vessels.

Previous reports found association of IA with SNP in both the 5’ and 3’ regions of SOX17. The SNP in the 2 regions were not in linkage disequilibrium and appear to have independent effects. The association with both regions of SOX17 was found in samples of European descent; however, a Japanese cohort demonstrated association only with the 5’ SNP. A subsequent study found that the association with the 5’ SNP varied among different populations, whereas the 3’ association remained robust in all but the Japanese cohorts. We were able to detect association only with SNP in the 3’ end of SOX17.

Smoking effects on IA may be moderated in part via genes that contain high-risk genetic variants. We found no evidence of an interaction between the most significant SNP in ANRIL and smoking. The data were consistent with the multiplicative effect of 2 variables in a logistic regression model in which the predicted risk is obtained by multiplying the 2 odds ratios. Thus, subjects who have the SNP risk allele and who also smoke have a greatly increased risk of IA as compared with subjects who have only the SNP risk allele. However, smoking remains the dominant risk factor and, because genotype cannot be altered, cessation of smoking would dramatically decrease the risk of IA, even among those having inherited a high-risk variant.

We did not replicate the other SNP found associated with IA in previous genome-wide association studies. Insufficient power may be an explanation for the failure to detect such an association, compared with the replication of associations on chromosome 8 and 9. Another possible reason may be that the associated risk is heterogeneous, and these risk variants were less important in our samples.

We used 2 discovery samples to test the hypothesis that unique risk variants segregate in familial IA. The most promising association from the familial IA analysis was with PDE1A. The same SNP were not genotyped in ARIC so we imputed them. However, there was no evidence of association with either of these SNP (rs1897472: $P=0.24$; rs3769801: $P=0.13$). This lack of replication suggests that this finding may have been a false-positive association.

This study had several important strengths. All case subjects underwent a uniform and rigorous evaluation that included review of medical records. The case subjects from DS1 came from densely affected, multiplex families who underwent a thorough evaluation of all putative cases, even when biological samples were not available. The control subjects, although recruited through multiple studies, also underwent a careful review to identify any evidence of a previous stroke or IA. Detailed data regarding smoking and hypertension, key risk factors for IA, were obtained from case subjects and control subjects, allowing us to evaluate whether our associations were independent of these critical covariates. This study also had some weaknesses. DS1 is of modest size and best-suited to detect risk factors of moderate effect. DS2 was larger, but still only modestly powered. Thus, both samples were best-suited to replicate previous findings rather than primary gene discovery. Furthermore, we used case subjects and control subjects from a variety of studies to achieve the critical sample size necessary to obtain genome-wide significance, which may have introduced genetic heterogeneity.

In summary, we provide further evidence that the association on chromosome 9p is attributable to variants in CDKN2BAS rather than CDKN2A or CDKN2B as initially reported. Furthermore, smoking increases the effect of the risk allele on chromosomes 8 and 9, and cessation of smoking.

### Table 3. Logistic Regression Model to Evaluate the Relationship Between Single Nucleotide Polymorphism Genotype and Smoking on the Relative Risk of Intracranial Aneurysm

<table>
<thead>
<tr>
<th>SNP rs1072737 (Chromosome 8)</th>
<th>Sample</th>
<th>OR</th>
<th>95% CI</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample 1</td>
<td>1.39</td>
<td>1.06 to 1.81</td>
<td>0.017</td>
<td></td>
</tr>
<tr>
<td>Sample 2</td>
<td>1.22</td>
<td>1.07 to 1.39</td>
<td>0.002</td>
<td></td>
</tr>
<tr>
<td>Meta-analysis</td>
<td>1.25</td>
<td>1.11 to 1.40</td>
<td>&lt;0.001</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SNP rs6475606 (Chromosome 9)</th>
<th>Sample</th>
<th>OR</th>
<th>95% CI</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample 1</td>
<td>1.52</td>
<td>1.18 to 1.95</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>Sample 2</td>
<td>1.33</td>
<td>1.18 to 1.50</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>Meta-analysis</td>
<td>1.36</td>
<td>1.22 to 1.52</td>
<td>&lt;0.001</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Smoking*</th>
<th>OR</th>
<th>95% CI</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample 1</td>
<td>5.34</td>
<td>3.57 to 7.95</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Sample 2</td>
<td>2.12</td>
<td>1.76 to 2.56</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Meta-analysis</td>
<td>2.51</td>
<td>2.14 to 2.96</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

CI indicates confidence interval; OR, odds ratio; SNP, single nucleotide polymorphism.

*Smoking refers to 20 pack-years of smoking.
may dramatically decrease the risk of IA, even among those who inherited a gene variant associated with small to moderate risk.

**Acknowledgments**

The authors thank the subjects and their families for participating in this research study.

**Sources of Funding**

This project was supported by R01NS39512, R01NS36695, K02NS060892, HHMI P0012375, HHSN268201100005C, HHSN 268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, HHSN268201100012C, R01HL087641, R01HL59367 R01HL086694, R01HL087641, R01HL59367 R01HL086694, HHSN268201100005C, HHSN 268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, HHSN268201100012C, R01HL087641, R01HL59367 R01HL086694, U01HG004402, HHSN268200625226C N01HC55015, N01HC55016, N01HC55018, N01HC55019, N01HC55020, N01HC55021, N01HC 55022, R01HL087641, U1LR0255005, U01 HL096917, and R01HL093029, and the Intramural Research Program of NGHRI (NIH). Addition funding was provided by the National Health and Medical Research Council of Australia and the Health Research Council of New Zealand.

**Disclosures**

None.

**References**


Genome-Wide Association Study of Intracranial Aneurysms Confirms Role of Anril and SOX17 in Disease Risk


Stroke. published online September 6, 2012;

Stroke is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2012 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/early/2012/09/06/STROKEAHA.112.656397

Data Supplement (unedited) at:
http://stroke.ahajournals.org/content/suppl/2012/09/06/STROKEAHA.112.656397.DC1

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/
SUPPLEMENTAL METHODS

Discovery Sample 1
The initial discovery sample consisted of individuals recruited through the Familial Intracranial Aneurysm (FIA) I study, in which familial cases that met one of several criteria involving multiple family members with IA, which would make the family appropriate for linkage analysis, were recruited through 26 clinical centers (41 sites) in North America, New Zealand, and Australia. Exclusion criteria included: (i) a fusiform-shaped unruptured IA of a major intracranial trunk artery; (ii) an IA which is part of an arteriovenous malformation; (iii) a family or personal history of polycystic kidney disease, Ehlers Danlos syndrome, Marfan’s syndrome, fibromuscular dysplasia, or Moya-Moya disease; or (iv) failure to obtain informed consent from the patient or family members. All medical records and relevant accompanying data were reviewed by a Verification Committee. For the present analysis, only individuals having an IA based on an intra-arterial angiogram, operative report, autopsy, or size ≥7 mm on non-invasive imaging (MRA, CTA) were considered “definite” cases. A set of independent unrelated cases was obtained by selecting one individual with definite IA from each FIA I family self-reported as Caucasian (n=389). The FIA study was approved by the Institutional Review Boards/Ethics Committees at all clinical and analytical centers and recruitment sites.

Controls for the Discovery Sample 1 were obtained from two population-based studies. The first was the NINDS-funded case-control Genetic and Environmental Risk Factors for Hemorrhage Stroke (GERFHS) study, which was designed to identify the important environmental and genetic risk factors for IA-related SAH as well as for spontaneous intracerebral hemorrhage. Controls identified by random-digit telephone dialing from the Greater Cincinnati/Northern Kentucky community and matched to enrolled cases by age (±5 years), gender, and race, had the same interview questions regarding environmental risk factors as FIA study participants. A set of 113 unrelated, Caucasian controls were selected for genotyping. In addition, 290 Caucasian controls free of stroke and IA were selected from the Cincinnati Control Cohort. The subjects in this cohort were identified by random-digit dialing from the Greater Cincinnati region during 2006. These subjects had blood drawn for DNA extraction as well as extensive interviews including detailed environmental exposures as well as detailed medical history of every major disease. Both studies were approved by the Institutional Review Boards of the University of Cincinnati and all participating hospitals.

Results from a subset of the cases (n=343) and controls (n=374) from Discovery Sample 1 were previously reported in Deka et al, 2010.

Discovery Sample 2
During FIA II study recruitment, the requirement for family history of IA was removed and both familial and sporadic IA cases were enrolled. The same exclusion
criteria were in place and all cases underwent the same rigorous review from the Verification Committee. A set of 829 Caucasian IA cases was selected for genotyping from this sample, and an additional 61 Caucasian sporadic aneurysmal SAH cases from the Greater Cincinnati/Northern Kentucky region were obtained from GERFHS.

This sample was augmented by Caucasian cases and controls identified from other studies, including those from the Australasian Cooperative Research on Subarachnoid hemorrhage Study (ACROSS), which was a prospective, population-based, case-control study of SAH undertaken in three cities in Australia and one city in New Zealand during the mid-1990s.\(^4\) ACROSS included incidence cases of SAH secondary to documented or presumed ruptured IA who were frequency-matched (by sex, 10-year age strata, and city of residence) to controls selected from electoral rolls in each city. Detailed information about key exposures, such as smoking, hypertension, family history of stroke/IA, was obtained by standardized interviews with subjects (or proxies) and where possible, blood samples were obtained for storage and future DNA extraction. Samples from a total of 160 cases and 168 controls were available for genotyping. This study was approved by the institutional review committees at 10 sites.

In addition, IA cases were recruited from a prospective cohort study of adult patients with spontaneous SAH due to IA confirmed by non-contrast CT and cerebral angiogram who were admitted to a tertiary-care referral center in San Francisco during 2003 to 2008. Additional FIA exclusion criteria were also applied to yield 184 samples from Caucasian subjects with detailed medical histories and blood banked for DNA. This study was approved by the institutional review committee at University of California, San Francisco.

Genotypic data from a further set of 1148 white controls was obtained through a collaborative agreement with the Atherosclerosis Risk in Communities (ARIC) study. In the ARIC sample, a subset of subjects who never had a stroke or TIA was matched to the Discovery Sample 2 cases by sex and, where possible, by age (±5 years). However, because the age of the ARIC controls was limited to 44–66, cases younger than 39 or older than 71 at onset were matched to controls outside of the 5-year criterion. Genotyping had been previously performed using the Affymetrix 6.0 array.\(^5\)

**Family history**

Positive family history for cases in Discovery Sample 1 was validated (supplemental text). For most Discovery Sample 2 cases and for controls in Discovery Sample 1, positive family history was based only on self-report by the subject. Family history was considered positive if any relative (not necessarily first-degree relative) was reported to have had a ruptured or unruptured IA. Family history of IA was not collected in controls for the ARIC or ACROSS studies.

**Genotyping and Quality Review**

Genotyping was performed using the Axiom array at the Affymetrix core labs for all samples except for the ARIC controls. Twenty-five internal samples were genotyped twice for quality control. This yielded a total of 2,219 samples sent for genotyping. However, only 2,140 samples with a QC (dQC) value ≥0.82 and an initial call rate of 97% were released. All released genotypes underwent a common quality review pipeline which included identification of sample duplicates, related individuals, and
gender discrepancies, which resulted in the removal of 59 samples. Prior to performing imputation, SNPs were excluded if there was: (i) improper mapping to Genome Reference Consortium GRCh37; (ii) a minor allele frequency (MAF) <0.03; (ii) a SNP call <95%; (iv) a Hardy Weinberg Equilibrium (HWE) p-value in controls of $p<10^{-2}$ and $p<10^{-4}$ in cases. From the 567,096 SNPs on the Axiom array, 473,238 were retained following this quality review.

A principal component analysis (PCA) was performed using Eigenstrat$^6$ and data from 11 HapMap phase III populations to identify clusters using the first two eigenvectors computed using the SNPs typed on both platforms. Samples clustering with the European American (CEU) reference set (PC1: 0.0024–0.0078; PC2: 0.0007–0.0049) were retained, and those outside this cluster which were likely to contain African, Asian, or Hispanic admixture were removed from further analysis (n=47 of the Axiom-genotyped samples); 16 non-European American samples from the ARIC set were also removed. Coordinates of study subjects for PCs 1 and 2 are illustrated in Supplemental Figure 2.

Genotypic data for the ARIC samples was obtained from the Affy 6.0 array.$^5$ These data also underwent quality review and SNPs were removed based on the same criteria listed above. From the 793,799 SNPs on the Affy 6.0 array that were provided by ARIC following their initial data review, a total of 619,514 were retained for analysis in this study.

### Power for the Samples

Discovery Sample 1 had low power to detect an allelic association at the genome-wide significance threshold ($5 \times 10^{-8}$), but had 50% power to detect an association with an odds ratio of 1.5 at the $10^{-4}$ screening threshold for common SNPs (MAF≥0.4). At the threshold for genome-wide significant evidence of association, Discovery Sample 2 had 80% or greater power to detect an odds ratios of 1.4 or greater for common SNPs (MAF≥0.4). Similarly, the meta-analysis sample had 90% or greater power to detect odds ratios of 1.4 or greater with common SNPs (MAF≥0.4) at the genome-wide significance level.

### Imputation

Imputation was performed for all autosomes using IMPUTE2 ([https://mathgen.stats.ox.ac.uk/impute/impute_v2.html](https://mathgen.stats.ox.ac.uk/impute/impute_v2.html)). All distinct samples genotyped on the Axiom array (n=2,115) were imputed together using the 1000Genomes haplotypes (n=1094; data freeze from Nov. 2010, Jun. 2011 phased haplotype release, mapped to GRCh37) as the phased reference panel. IMPUTE2 implements a method that can utilize an additional reference panel data as well ([https://mathgen.stats.ox.ac.uk/impute/using_multi_population_reference_panels.html](https://mathgen.stats.ox.ac.uk/impute/using_multi_population_reference_panels.html)); we included the 1,148 ARIC samples as another unphased reference panel to maximize available information at each imputed SNP.

Because Discovery Sample 2 was genotyped on two platforms, with nearly complete confounding by type of sample (i.e., all cases on Axiom; the majority of controls on Affy 6.0), extensive and detailed quality review was performed to ensure that spurious association was not detected based on platform effects. As suggested by
we reviewed several SNP metrics, including imputation quality (information) and differences in SNP minor allele frequency in controls genotyped on the Axiom platform, and the ARIC controls genotyped on the Affy 6.0. We removed all SNPs with low to moderate imputation quality (information score <0.50) as well as those SNPs with a significant difference in minor allele frequency between the two sources of control samples (p<0.1). To further reduce the influence of rare SNPs, which would typically have less accurate imputation, we removed all SNPs with a minor allele frequency less than 5%. Using this aggressive filtering approach, we retained 453,699 SNPs for analysis of Discovery Sample 2. Remaining uncertainty in the imputed genotypes after application of the aggressive information score and minor allele frequency filters was modeled using the “-method score” option in IMPUTE2. We would expect a slight loss of power in the association tests due to the uncertainty in genotypes; however, previous studies indicate this power loss is minimal, on the order of 7% of the effective sample size on average.

A few critical SNPs were not genotyped in the ARIC sample. Therefore, we performed imputation of specific SNPs using IMPUTE2. The same 1000G haplotypes are used as the phased reference samples. The 2115 samples genotyped on the Axiom array were used as the unphased reference panel.

### Gene x Smoking Relationship

Since cigarette smoking is a very strong risk factor in IA, we examined the possible interaction of the most highly associated SNPs on chromosomes 8 and 9 with cigarette smoking. On chromosome 8, we used a SNP (rs1072737) that was imputed in the samples genotyped on the Axiom array. The imputation procedure generates each individual’s probability of each genotype for this SNP. To avoid ambiguity, in cases where the probability of one SNP genotype was greater than 80%, we assigned the individual that genotype. However, if all genotypic probabilities were less than 80%, then that individual was omitted from the analysis (9% of samples removed; n=285). On chromosome 9 we used SNP (rs6475606), which was genotyped on both the Axiom and Affy 6.0 arrays. Using the same cases and controls from discovery samples 1 and 2, a logistic regression model was employed to test for departures from a multiplicative relationship between the risk allele scores (no risk allele=0, 1 risk allele=1, and 2 risk alleles=2) and cumulative exposure to smoking as measured by pack-years. Since the distribution of pack-years is highly skewed, we used the log of pack-years in the logistic model, with 0.05 pack-years assigned to the never-smokers. The logistic models were fitted for discovery samples 1 and 2, and each model was adjusted for age and sex. We combined estimates using meta-analysis with individual sample results weighted by the inverse variance of the sample estimates.

### ADDITIONAL INTERPRETATION FOR TABLE 3

The OR for any given number of K pack-years can be calculated using the following equation: $OR = \exp(\beta(\ln(K) + 2.99))$ where $\beta$ = the regression coefficient for log(pkyrs). $\beta=0.155$ for rs6475606 and $\beta=0.154$ for rs1072737. For example, the odds ratio for 40 pack years of smoking for subjects in the model of the rs6475606 risk allele = $\exp(0.155(\ln(40) + 2.99)) = 2.82$. To determine the odds ratio for presence of two risk
alleles of rs6475606 (homozygous state) and 40 pack years of smoking, one would multiply \((1.36)^2 = 1.85\) (two risk alleles) \(x 2.82\) which equals an odds ratio of 5.22.
Supplemental Figure 1A
Supplemental Figure 1B
Supplemental Figure 1D
Supplemental Figure 1E
Supplemental Figure 1F
Figure 1: Comparison with previously reported results from genomewide association studies.\textsuperscript{9,10} (A) Chromosome 4q31.23 \textit{(EDNRA)}; (B) Chromosome 10q24.32 \textit{(CNNM2)}; (C) Chromosome 12q22; (D) Chromosome 13q13.1 \textit{(KL/STARD13)}; (E) Chromosome 18q11.2 \textit{(RBBP8)}; (F) Chromosome 20p12.1. When the most significant SNP from the initial report was available in the meta analysis, a white diamond indicates the SNP, and the SNP is listed at the top of the figure. If the SNP was not available in the meta analysis sample, an arrow is used to denote the position of that SNP within this map of markers. Each circle symbol within the graph indicates the p-value for a SNP at that position in the meta analysis. The color of the square symbol denotes the extent of linkage disequilibrium (as computed by $r^2$) with the SNP reported in the initial report.
**Figure 2: Principal component clustering plot for genotyped study subjects.**

Genotyped individuals are shown for PC1 (x-axis) and PC2 (y-axis). Reference populations YRI (African) and CHB/JPT (Asian) are shown as green and red symbols, respectively, with study subjects as orange circles. As described in the text, genotyped subjects clustering outside the area defined by the CEU (European-American) reference sample (PC1: 0.0024–0.0078; PC2: 0.0007–0.0049) were identified and excluded from association analyses.
SUPPLEMENTAL REFERENCES


4. Anderson C, Ni Mhurchu C, Scott D, Bennett D, Jamrozik K, Hankey G. Triggers of subarachnoid hemorrhage: Role of physical exertion, smoking, and alcohol in the australasian cooperative research on subarachnoid hemorrhage study (across). *Stroke*. 2003;34:1771-1776


7. Sinnott JA, Kraft P. Artifact due to differential error when cases and controls are imputed from different platforms. *Hum Genet*. 2012;131:111-119


Clinical Centers – University of Alabama at Birmingham: W. Fisher (PI), H. Forson (coordinator); Clinical Trials Research Unit, University of Auckland and Auckland City Hospital, New Zealand: C. Anderson (PI), E. Mee (PI), C. Howe (coordinator), S. Vos (coordinator); Royal Perth Hospital, Sir Charles Gairdner Hospital, Royal Adelaide Hospital, Royal Melbourne Hospital, Alfred Hospital, Westmead Hospital, Royal North Shore Hospital, Royal Prince Alfred Hospital, Australia: C. Anderson (PI), G. Hankey (PI), N. Knuckey (PI), J. Laidlaw (PI), P. Reilly (PI), N. Dorsch (PI), M. Morgan (PI), M. Besser (PI), J. Rosenfeld (PI), K. Athanasiadis (coordinator), A. Claxton (coordinator), V. Dunne (coordinator), J. Griffith (coordinator), J Davidson (coordinator), S. Pope (coordinator), Amanda Froelich (coordinator); Brigham & Women’s Hospital: A. Day (PI), R. Brach (coordinator); University of Cincinnati: D. Woo (co-PI), M. Zuccarello (co-PI), A. Ringer (co-PI), H. Yeh (co-PI), K. Franklin (coordinator); Cleveland Clinic Foundation: P. Ramussen (PI), D. Andrews-Hinders (coordinator), T. Wheeler (coordinator); Columbia University: E. S. Connolly (PI), R. Sacco (co-PI), D. LaMonica (coordinator); University of Florida: S. B. Lewis (PI), A. Royster (coordinator); Indianapolis Neurosurgical Group: T. Payner (PI), N. Miracle (coordinator); Johns Hopkins: K. Murphy (PI), B. Kohler (coordinator); Massachusetts General Hospital: C. Ogilvy (PI), D. Buckley (coordinator), J. Manansala (coordinator); London Health Science Center Research Inc.: G. Ferguson (PI), C. Mayer (coordinator), J. Peacock (coordinator); Notre Dame Hospital: G. Rouleau (PI), A. Desjarlais (coordinator); University of Maryland: E. F. Aldrich (PI), C. Aldrich (coordinator), C. Byard (coordinator); Mayo Clinic: R. D. Brown (PI), L. Jaeger (coordinator); University of Michigan: L. Morgenstern (PI), M. Concannon (coordinator); New Jersey Medical School: A. I. Qureshi (PI), P. Harris-Lane (coordinator); Northwestern University: H. Batjer (PI), G. Joven (coordinator), S. Thompson (coordinator); University of Ottawa: M. T. Richard (PI), A. Hopper (PI); University of Pittsburgh: A. B. Kassam (PI), K. Lee (coordinator); University of California, San Francisco: C. Johnston (PI), K. Katsura (coordinator); University of Southern California: S. Giannotta (PI), D. Fishback (coordinator); Stanford University Medical Center: G. Steinberg (PI), D. Luu (coordinator), M. Coburn (coordinator); University of Texas at Houston: M. Malkoff (PI), A. Wojner (coordinator); University of Virginia: N. Kassel (PI), B. Worrall (co-PI), G.
Radakovic (coordinator); University of Washington: D. Tirschwell (PI), P. Tanzi (coordinator); Washington University: C. Derdeyn (PI), M. Catanzaro (coordinator); University of Manitoba: A. Kaufmann (PI), D. Gladish (coordinator).

**Executive Committee:** Craig Anderson, MD; Joan E. Bailey-Wilson, PhD; Joseph P. Broderick, MD; Robert Brown, Jr. MD; E. Sander Connelly, MD; Ranjan Deka, PhD; Tatiana Foroud, PhD; John Huston, MD; Nerissa Ko, MD; Carl Langefeld, PhD; Irene Meissner, MD; Guy Rouleau MD, PhD; Laura Sauerbeck, RN, MS; Bradford Worrall, MD; Daniel Woo, MD, MS.

**Imaging Phenotyping Committee:** John Huston, MD; David Kallmes, MD; Harry Cloft, MD.

**Medical Record Phenotyping:** Joseph P. Broderick, MD; Robert Brown, Jr. MD; Daniel Woo, MD, MS; Irene Meissner, MD; David Wiebers, MD; Dawn Kleindorfer, MD; Matthew L. Flaherty, MD; Jason Mackey, MD, MS; Shannon Kohake, MD; Sharyl Martini, MD, PhD; Felipe De Los Rios La Rosa, MD.