ARIC Manuscript Proposal #2242

PC Reviewed: 10/8/13	**Status: A**	**Priority: 2**
SC Reviewed: _____	**Status: _____**	**Priority: ____**

1.a. Full Title: Genomewide Association of Intercranial Aneurysm Identifies a New Association on Chromosome 7

b. Abbreviated Title (Length 26 characters): GWAS in Intracranial Aneurysms

2. Writing Group:
Writing group members:

Indiana University: Tatiana Foroud, Dongbing Lai, Daniel Koller
University of Cincinnati: Joseph Broderick, Charles Moomaw
Analysts from each of the groups contributing data.

I, the first author, confirm that all the coauthors have given their approval for this manuscript proposal. __TF__ **[please confirm with your initials electronically or in writing]**

First author: **Tatiana Foroud**
Address: 410 W. 10th Street
Indianapolis, IN 46202

Phone: 317-278-1291 Fax: 317-278-1100
E-mail: tforoud@iu.edu

ARIC author to be contacted if there are questions about the manuscript and the first author does not respond or cannot be located (this must be an ARIC investigator).

Name: **Thomas H. Mosley**
Address: University of Mississippi Medical Center
2500 N. State St.
Jackson, MS 39216-4505

Phone: (601) 984-2763 Fax: (601) 815-3422
E-mail: tmosley@umc.edu

3. Timeline: Anticipate submission by November 1, 2013

4. Rationale:

Common variants contribute to the risk of intracranial aneurysm. Using a meta-analytic approach, data from multiple case control studies will be combined to maximize the power to identify novel genes contributing to IA susceptibility. This is an expansion of a previously approved request to utilize ARIC data as part of a meta-analysis (Ancillary study 2009.25).

5. Main Hypothesis/Study Questions:

Common variants contribute to the risk of intracranial aneurysm. Using a meta-analytic approach, data from multiple case control studies will be combined to maximize the power to identify novel genes contributing to IA susceptibility. This is an expansion of a previously approved request to utilize ARIC data as part of a meta-analysis.

6. Design and analysis (study design, inclusion/exclusion, outcome and other variables of interest with specific reference to the time of their collection, summary of data analysis, and any anticipated methodologic limitations or challenges if present).

Phenotype

Familial Intracranial Aneurysm Study (FIA Study): Families having at least 2 members with an intracranial aneurysm (IA) were ascertained 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. This sample has been included in a previous report as discovery sample I (Foroud et al, 2012).

Further recruitment was undertaken as part of the FIA Study and 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 has been included in a previous report as discovery sample II (Foroud et al, 2012). 680 Caucasian IA cases were genotyped as discovery sample 3A.

Australasian Cooperative Research on Subarachnoid Hemorrhage Study (ACROSS): 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. This sample has been has been included in a previous report (Foroud et al, 2012).

UCSF: 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. This sample has been has been included in a previous report (Foroud et al, 2012).

ARIC: Genotypic data from a 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). These data were from the Brain MRI Study (2004-2006). 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] This sample has been has been included in a previous report (Foroud et al, 2012).

Cincinnati Control Cohort and Genetic and Environmental Risk Factors for Hemorrhage Stroke: Controls 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. Another set of controls free of stroke and IA were selected from the Cincinnati Control Cohort (CCC). 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. 113 GERFHS and 290 CCC controls have been included in a previous report (Foroud et al, 2012). 375 GERFHS and 7 CCC controls were not previously analyzed.

Krakow, Poland: IA cases were recruited from patients of the Department of Neurology and the Department of Neurosurgery and Neurotraumatology of the Jagiellonian University in Krakow. Both subjects with ruptured IAs and with unruptured IA were recruited. Presence of IA was confirmed by intra-arterial angiogram, CTA, MRA or intraoperatively. A total of 504 IA patients were included. The control group included 514 unrelated subjects taken from the population of southern Poland. Control subjects had no apparent neurological disease based on the findings in a structured questionnaire and a neurological examination. All subjects were of European descent. Information about key demographics, family history and risk factors were obtained using a standardized questionnaire. The study was approved by the institutional review board of the Jagiellonian University.

Genotyping and Statistical Analysis

Genotyping and Quality Review

Genotyping of all samples except ARIC was performed using the Axiom array at the Affymetrix core labs. Genotyping was performed in four batches using similar methods and quality control. Forty-eight internal samples were genotyped twice for quality control. This yielded a total of 4,331 samples sent for genotyping. However, only 4,249 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 118 samples. Prior to performing imputation, SNPs were excluded if there were: (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 $<10^{-4}$. MAF, call rate and HWE p-values were calculated by combing all 4 batches together. From the 597,320 SNPs on the Axiom array, 464,632 were retained following this quality review.

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 autosome SNPs on the Affy 6.0 array that were provided by ARIC following their initial data review, a total of 626,645 were retained for imputation in this study.

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 were retained, and those outside this cluster which were likely to contain African, Asian, or Hispanic admixture were removed from further analysis (n=61 of the Axiom-genotyped samples); 16 non-European American samples from the ARIC set were also removed.

Imputation

Imputation was performed for all autosomes using IMPUTE2 (https://mathgen.stats.ox.ac.uk/impute/impute_v2.html). All distinct samples genotyped on the Axiom array (n=4060) were imputed together using the 1000Genomes haplotypes (n=1092; data freeze from Nov. 2010, May. 2011, March 2012 phased haplotype release) as the phased reference panel. Only variants with more than one minor copy across all 1000Genome populations were imputed. Original genotypes were not overwritten. ARIC sample (n=1132) were imputed separately using the same reference panel. 30,057, 018 and 30,061,894 SNPs were imputed from AXIOM array data and ARIC data, respectively. From which, 1,195,878 SNPs were typed in either AXIOM array or Affy 6.0, and these SNPs were used in analysis.

Statistical Analysis

Because our sample was genotyped on two platforms, with sample typed in Affy 6.0 are all controls, extensive and detailed quality review was performed to ensure that spurious association was not detected based on platform effects. As suggested by Sinnott and Kraft (Sinnott and Kraft, 2012) 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 imputation quality (information score <0.30) 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 672,213 SNPs for analysis. 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 SNPTEST V2. 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 (Nair et al, 2009). All samples were analyzed together with genomic control applied to correct for inflation.

The genotyped and imputed SNPs were used to test for association with IA susceptibility using a logistic regression model. No additional covariates were necessary. Analysis was performed using the SNPTEST

v2 software. All analyses were performed with an additive model of SNP effect. Imputed genotypes were encoded in the logistic model as the expected allele count.[14] For autosome SNPs, all samples were analyzed together. For chromosome X, only SNPs that were on the AXIOM array were included. The ARIC data only included genotypes for the autosomes and so were not included in analyses of the X chromosome. Using the same QC as above, 13,070 chromosome X SNPs were used. Genomic control was applied to correct for inflation. Our dataset included only 685,283 SNPs for the association analysis. Therefore, we applied a Bonferroni correction to obtain the appropriate genome-wide threshold for significance ($0.05/685,283 = p<7.3 \times 10^{-8}$).

To test whether there might be more than one risk variant in a particular gene or gene region contributing to the association, we performed conditional analyses. For the region on chromosome 7, we identified the SNP with the most extreme p-value in the meta-analysis. We then modified the logistic regression model to include the genotype at the most significant SNP in the region and test for association with other SNPs in the region. Finally, we reviewed the p-value for the other SNPs in the region to determine whether any other SNPs remained statistically significant.

7.a. Will the data be used for non-CVD analysis in this manuscript? ____ Yes _X___ No

b. If Yes, is the author aware that the file ICTDER03 must be used to exclude persons with a value RES_OTH = "CVD Research" for non-DNA analysis, and for DNA analysis RES_DNA = "CVD Research" would be used? ____ Yes ____ No
(This file ICTDER has been distributed to ARIC PIs, and contains the responses to consent updates related to stored sample use for research.)

8.a. Will the DNA data be used in this manuscript?
__X_ Yes ____ No

8.b. If yes, is the author aware that either DNA data distributed by the Coordinating Center must be used, or the file ICTDER03 must be used to exclude those with value RES_DNA = "No use/storage DNA"?
__X__ Yes ____ No

9. The lead author of this manuscript proposal has reviewed the list of existing ARIC Study manuscript proposals and has found no overlap between this proposal and previously approved manuscript proposals either published or still in active status. ARIC Investigators have access to the publications lists under the Study Members Area of the web site at: http://www.cscc.unc.edu/ARIC/search.php

___X__ Yes _____ No

10. What are the most related manuscript proposals in ARIC (authors are encouraged to contact lead authors of these proposals for comments on the new proposal or collaboration)?
N/A

11.a. Is this manuscript proposal associated with any ARIC ancillary studies or use any ancillary study data? __X__ Yes __ _ No

11.b. If yes, is the proposal

_X_ A. primarily the result of an ancillary study (list number: Ancillary study #2009.25 and ARIC Brain MRI: 1999.01**)**

___ B. primarily based on ARIC data with ancillary data playing a minor role (usually control variables; list number(s)* _____ _____)

*ancillary studies are listed by number at http://www.cscc.unc.edu/aric/forms/

12a. Manuscript preparation is expected to be completed in one to three years. If a manuscript is not submitted for ARIC review at the end of the 3-years from the date of the approval, the manuscript proposal will expire.

12b. The NIH instituted a Public Access Policy in April, 2008 which ensures that the public has access to the published results of NIH funded research. It is **your responsibility to upload manuscripts to PUBMED Central** whenever the journal does not and be in compliance with this policy. Four files about the public access policy from http://publicaccess.nih.gov/ are posted in http://www.cscc.unc.edu/aric/index.php, under Publications, Policies & Forms. http://publicaccess.nih.gov/submit_process_journals.htm shows you which journals automatically upload articles to Pubmed central.