Rare Coding Variation and Risk of Intracerebral Hemorrhage

Farid Radmanesh, MD, MPH; Guido J. Falcone, MD, ScD; Christopher D. Anderson, MD, MMSc; David McWilliams, PhD; William J. Devan, BS; W. Mark Brown, MA; Thomas W.K. Battey, BS; Alison M. Ayres, BA; Miriam R. Raffeld, BS; Kristin Schwab, BA; Guangyun Sun, PhD; Ranjan Deka, PhD; Anand Viswanathan, MD, PhD; Joshua N. Goldstein, MD, PhD; Steven M. Greenberg, MD, PhD; David L. Tirschwell, MD, MSc; Scott L. Silliman, MD; Magdy Selim, MD, PhD; James F. Meschia, MD; Devin L. Brown, MD, MS; Bradford B. Worrall, MD, MSc; Carl D. Langefeld, PhD; Daniel Woo, MD; Jonathan Rosand, MD, MSc

Background and Purpose—Intracerebral hemorrhage has a substantial genetic component. We performed a preliminary search for rare coding variants associated with intracerebral hemorrhage.

Methods—A total of 757 cases and 795 controls were genotyped using the Illumina HumanExome Beadchip (Illumina, Inc, San Diego, CA). Meta-analyses of single-variant and gene-based association were computed.

Results—No rare coding variants were associated with intracerebral hemorrhage. Three common variants on chromosome 19q13 at an established susceptibility locus, encompassing TOMM40, APOE, and APOC1, met genome-wide significance (P<5e−08). After adjusting for the APOE epsilon alleles, this locus was no longer convincingly associated with intracerebral hemorrhage. No gene reached genome-wide significance level in gene-based association testing.

Conclusions—Although no coding variants of large effect were detected, this study further underscores a major challenge for the study of genetic susceptibility loci; large sample sizes are required for sufficient power except for loci with large effects. (Stroke. 2015;46:00-00. DOI: 10.1161/STROKEAHA.115.009838.)

Key Words: apolipoproteins E • cerebral hemorrhage • genome-wide association study

Genetic variation plays a substantial role in the risk of intracerebral hemorrhage (ICH).1 Genome-wide association studies have identified commonly variants associated with risk of ICH, both lobar and nonlobar subtypes.2 The degree to which rare genetic variants, those with minor allele frequencies far smaller than those of variants typically discovered through genome-wide association studies, contribute to this risk is unknown. Preliminary targeted sequencing studies have supported a possible role for rare variants in sporadic ICH.3 Recently, the exome array has emerged as an efficient, cost-effective tool to bridge array-based common variant association studies and whole-exome or whole-genome sequencing to identify coding variation underlying common conditions. The goal of this study was to explore the role of exonic variants in risk of ICH, using exome array.

Methods

Study subjects, genotyping, and quality control procedures are described in the Methods in the online-only Data Supplement.
Scores and minor allele frequency (MAF) for individual variants and a covariance matrix for each gene were computed, including age, sex, and the first 2 principal components as covariates in the model. Inverse variance–weighted meta-analysis of score tests was computed for both common and rare variants.

As MAF decreases, single-variant analysis loses the power to reach genome-wide significance, even in the presence of a true association. Therefore, variants within each gene or region of interest are aggregated to increase the power to detect variants with small effects. We applied sequence kernel association test (SKAT), SKAT-O, and T1 count tests for gene-based analysis. In analysis using SKAT, each single nucleotide polymorphism was weighted by the inverse of its SE and its MAF, where variants with lower MAF are relatively upweighted. In the T1 count test, each variant was weighted equally, irrespective of their MAF. The association models were adjusted for age, sex, and the first 2 principal components.

We performed association analysis in all subjects, as well as separately for lobar ICH. Analysis of nonlobar ICH was not performed because of small sample size. Quality control was performed using PLINK v1.07. All other analyses were performed using seqMeta package in R version 3.0.2.

Results
After excluding subjects for quality (n=31) and genetic outliers (n=56), there were 1553 subjects for analysis (Table I in the online-only Data Supplement).

In single-variant analysis, we identified a susceptibility locus at chr19q13 (P<5e−08), including 3 common variants with MAF ranging from 13% to 19% (Table; Figure I in the online-only Data Supplement). The top variant at this locus was rs769449, which is an intronic single nucleotide polymorphism on APOE (P=1.94e−11; odds ratio, 1.97 [95% confidence interval, 1.62–2.40]). There was no evidence of heterogeneity across 2 studies (Figure 1). These variants are in moderate linkage disequilibrium, with r² estimates ranging 0.4 to 0.6 (Figure 2).

The 19q13 locus encompasses TOMM40, APOE, and APOC1. Common variants in this locus have been associated with several traits, including lipid levels, Alzheimer disease, cerebral amyloid angiopathy, and ICH. Given the association of APOE ε2 and ε4 alleles with ICH, we adjusted for these alleles, which had been previously genotyped in the majority of study subjects (Table II in the online-only Data Supplement). This adjustment resulted in loss of the observed signal, suggesting that these associations arose from the effect of ε2 and ε4 alleles (Table).

No low frequency variant or gene emerged as associated with ICH or the lobar subtype using SKAT, SKAT-O, or burden tests before and after adjustment for the ε2 and ε4 alleles. The strongest association for the gene-based analyses was observed for GADLI in the T1 count test after adjustment for the epsilon alleles (P=6.37e−05; cumulative MAF=3.3%).

Discussion
Common genetic variation seems to play a substantial role in ICH risk and key clinical features, including clinical outcome. Ongoing genome-wide association studies are designed to detect common variants, but they may miss rare variation. The contribution from rare variation is less substantiated.

The present analysis did not identify any rare coding variants for ICH. Our effort to identify coding variation with modest effect sizes was limited by inadequate statistical power. The gene-based tests can partially compensate for this limitation, but still lack of sufficient number of observations of rare variants in such small sample sizes prohibits taking full advantage of this approach. We estimated that our power was ≈7% for
detection of a significant association at $P<1\text{e}^{-06}$ (corrected for multiple tests in the gene-based analysis) at maximum odds ratio=5 when MAF=0.0001.\textsuperscript{9} Our data therefore suggest that most genetic risk for ICH resides within common and rare variants with modest effect size. Accurate estimation of the extent to which rare variants contribute to risk of ICH will require larger scale sequencing studies with coverage of both common and rare variants.

The development of international consortia has facilitated recruitment of hundreds of thousands of subjects with common diseases such as ischemic stroke and accelerated the rate of genetic discoveries for complex traits. With decreasing costs of sequencing studies and further expansion of consortia, genetic characterization of less common conditions such as ICH will become more feasible.

**Sources of Funding**

This study was supported by grants R01NS073344, R01NS059727, and 5K23NS059774 from the National Institutes of Health–National Institute of Neurological Disorders and Stroke (NIH-NINDS) and 0755984T from the American Heart Association. Computation support was provided, in part, by the Wake Forest School of Medicine Institute for Heart, Vascular and Stroke Care. Dr Worrall received a research grant from the NIH. Dr Rosand received research grants from the NIH.

**Disclosures**

Dr Worrall is the Associate Editor for the journal Neurology. Dr Rosand is a consultant to Boehringer Ingelheim. The other authors report no conflicts.

**Table.** Variants Associated With Intracerebral Hemorrhage at $P<\text{e}^{-08}$ in Meta-Analysis

<table>
<thead>
<tr>
<th>CHR</th>
<th>SNP</th>
<th>Gene</th>
<th>BPP</th>
<th>Coded Allele</th>
<th>MAF</th>
<th>Unadjusted for $\epsilon$ Alleles</th>
<th>Adjusted for $\epsilon$ Alleles</th>
</tr>
</thead>
<tbody>
<tr>
<td>19</td>
<td>rs769449</td>
<td>APOE</td>
<td>45410002</td>
<td>A</td>
<td>0.13</td>
<td>1.97 [1.62–2.40]</td>
<td>1.94e-11</td>
</tr>
<tr>
<td>19</td>
<td>rs2075650</td>
<td>TOMM40</td>
<td>45395619</td>
<td>G</td>
<td>0.15</td>
<td>1.70 [1.42–2.03]</td>
<td>2.14e-08</td>
</tr>
<tr>
<td>19</td>
<td>rs4420638</td>
<td>APOC1</td>
<td>45422946</td>
<td>G</td>
<td>0.19</td>
<td>1.62 [1.38–1.89]</td>
<td>2.54e-08</td>
</tr>
</tbody>
</table>

\(\epsilon\) indicates APOE epsilon alleles 2 and 4; BPP, base pair position; CI, confidence interval; CHR, chromosome; MAF, minor allele frequency; OR, odds ratio; and SNP, single nucleotide polymorphism.

**References**

Rare Coding Variation and Risk of Intracerebral Hemorrhage

Stroke. published online June 25, 2015;
Stroke is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2015 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/2015/06/25/STROKEAHA.115.009838

Data Supplement (unedited) at:
http://stroke.ahajournals.org/content/suppl/2015/06/25/STROKEAHA.115.009838.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

Study participants

We performed exome array genotyping of 1,639 individuals with intracerebral hemorrhage (ICH) enrolled in two prospective studies, including Genetics of Cerebral Hemorrhage with Anticoagulation (GOCHA)\(^1\) and Genetic and Environmental Risk Factors for Hemorrhagic Stroke (GERFHS).\(^2\) Primary ICH was defined as a new and acute neurological deficit with compatible brain imaging showing intraparenchymal bleeding. GOCHA included subjects older than 55 years, self-identified as descendants of European ancestry. GERFHS recruited individuals older than 18 years of age with European (94%) or African American (6%) ancestry.

Case subjects were further categorized as those with lobar ICH when the hematoma appeared to originate in the cerebral cortico-subcortical junction, and non-lobar ICH when hemorrhage was located in deep supratentorial or infratentorial locations. Subjects with multiple hemorrhages, primary intraventricular hemorrhage, and secondary causes of ICH including vascular malformations, aneurysms, neoplasms, trauma, and hemorrhagic transformation of acute infarction were excluded.

Controls were stroke-free subjects randomly selected from the same population, using clinic-based sampling (GOCHA) or random-digit dialing (GERFHS). The institutional review board and ethics committees of participating institutions approved these studies and written informed consent was obtained from all participants or their next of kin.

Genotyping and quality control

Genotyping was performed at the Broad Institute using Illumina HumanExome-12v1_A Beadchip, which includes 247,870 markers with a focus on non-synonymous, splice-site, and stop gain/loss variants. Genotype calling was completed using zCall algorithm.\(^3\)
Quality control was performed according to standard procedures. Samples were excluded with missing rate >2%, mismatch between the reported gender and gender inferred from the genotype data, and first- or second-degree relatives identified based on identity-by-descent allele sharing (pi-hat>0.185). Variants with call rate <95%, mean heterozygosity >±3 standard deviation difference from the mean, departure from Hardy-Weinberg equilibrium at p<1×10⁻⁶ in control subjects, or differential missingness in cases and controls were excluded. All SNPs were aligned on the forward strand and coded to the same minor alleles in both datasets.

We used ancestry-informative markers implemented in the exome array to perform principal component analysis using EIGENSTRAT. Genetic outliers were identified by visual inspection of principal component plots incorporating individuals from phase 3 HapMap as the reference. The majority of subjects clustered with the CEU (Northern Europeans from Utah) and TSI (Tuscans from Italy) populations. Because the small number of African American subjects in the GERFHS study prohibitively limited power to detect genetic associations, only European subjects were included for further analyses.
**Supplemental Table I. Characteristics of studies**

<table>
<thead>
<tr>
<th>Covariate</th>
<th>GOCHA(^1)</th>
<th>GERFHS(^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lobar (n = 432)</td>
<td>Control (n = 468)</td>
</tr>
<tr>
<td>Age, mean (SD)</td>
<td>74 (10)</td>
<td>73 (8)</td>
</tr>
<tr>
<td>Female, n (%)</td>
<td>233 (54)</td>
<td>231 (49)</td>
</tr>
<tr>
<td>Hypertension, n (%)</td>
<td>275 (64)</td>
<td>282 (60)</td>
</tr>
<tr>
<td>Warfarin, n (%)</td>
<td>69 (16)</td>
<td>69 (15)</td>
</tr>
<tr>
<td>Antiplatelets, n (%)</td>
<td>189 (44)</td>
<td>271 (58)</td>
</tr>
</tbody>
</table>

\(^1\) Genetics of Cerebral Hemorrhage with Anticoagulation study  
\(^2\) Genetic and Environmental Risk Factors for Hemorrhagic Stroke study
Supplemental Table II. *APOE* ε2 and ε4 allele frequency

<table>
<thead>
<tr>
<th></th>
<th>GOCHA¹</th>
<th></th>
<th></th>
<th>GERFHS²</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cases</td>
<td>Controls</td>
<td>p*</td>
<td>Cases</td>
<td>Controls</td>
</tr>
<tr>
<td>ε2</td>
<td>11.3</td>
<td>8.2</td>
<td>0.01</td>
<td>12.2</td>
<td>8.3</td>
</tr>
<tr>
<td>ε4</td>
<td>22.7</td>
<td>12.9</td>
<td>1.78e-06</td>
<td>22.8</td>
<td>12.1</td>
</tr>
</tbody>
</table>

1- Genetics of Cerebral Hemorrhage with Anticoagulation study
2- Genetic and Environmental Risk Factors for Hemorrhagic Stroke study
3- ε: APOE epsilon alleles

* Logistic regression model adjusting for age, gender, the first two principal components
References


Supplemental Figure I. Manhattan plot: Meta-analysis of single-variant association testing