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;46:2299-2301. 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 common 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.

Received April 23, 2015; final revision received April 23, 2015; accepted May 15, 2015.

From the Center for Human Genetic Research (F.R., G.J.F., C.D.A., W.J.D., T.W.K.B., M.R.R., J.R.), Division of Neurocritical Care and Emergency Neurology, Department of Neurology (F.R., G.J.F., C.D.A., W.J.D., T.W.K.B., M.R.R., J.R.), J. Philip Kistler Stroke Research Center (F.R., G.J.F., C.D.A., W.J.D., T.W.K.B., A.M.A., M.R.R., K.S., A.V., S.M.G., J.R.), and Department of Emergency Medicine (J.N.G.), Massachusetts General Hospital, Boston; Program in Medical and Population Genetics, Broad Institute, Cambridge, MA (F.R., G.J.F., C.D.A., W.J.D., T.W.K.B., M.R.R., J.R.); Department of Biostatistical Sciences, Center for Public Health Genomics, Wake Forest School of Medicine, Winston-Salem, NC (D.M., W.M.B., C.D.L.); Quinnipiac University, Frank H. Netter MD School of Medicine, Hamden, CT (W.J.D.); Department of Environmental Health (G.S., R.D.) and Department of Neurology and Rehabilitation (D.W.), University of Cincinnati College of Medicine, OH; Department of Neurology, Harborview Medical Center, University of Washington, Seattle (D.L.T.); Department of Neurology, University of Florida College of Medicine, Jacksonville (S.L.S.); Department of Neurology, Beth Israel Deaconess Medical Center, Boston, MA (M.S.); Department of Neurology, Mayo Clinic, Jacksonville, FL (J.F.M.); Department of Neurology, University of Michigan Health System, Ann Arbor (D.L.B.); and Departments of Neurology and Public Health Sciences, University of Virginia Health System, Charlottesville (B.B.W.).

Guest Editor for this article was Martin Dichgans, MD.

The online-only Data Supplement is available with this article at http://stroke.ahajournals.org/lookup/suppl/doi:10.1161/STROKEAHA.115.009838/-/DC1.

Correspondence to Jonathan Rosand, MD, MSc, Center for Human Genetic Research, Massachusetts General Hospital, 185 Cambridge St, CPZN 6810, Boston, MA 02114. E-mail jrosand@partners.org

© 2015 American Heart Association, Inc.

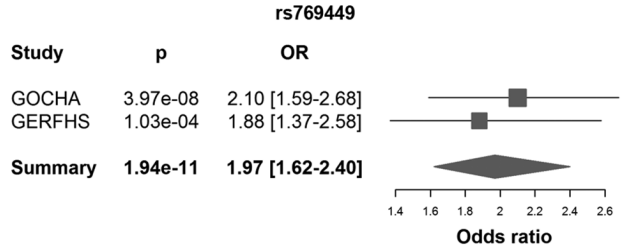

Figure 1. Forest plot depicting effect estimates for rs769449. OR indicates odds ratio; GERFHS, Genetic and Environmental Risk Factors for Hemorrhagic Stroke; and GOCHA, Genetics of Cerebral Hemorrhage With Anticoagulation.

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.[4] 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.[5]

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 r^2 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.[6,7] 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).[8] 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 *GADL1* 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.[1] 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

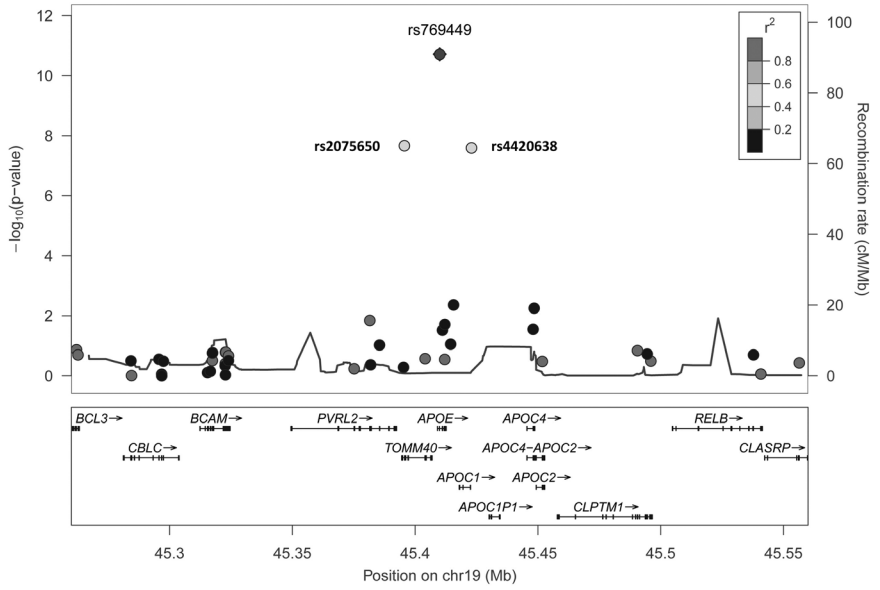

Figure 2. Regional association plot for 19q13 locus.

Table. Variants Associated With Intracerebral Hemorrhage at *P*<5e−08 in Meta-Analysis

CHR	SNP	Gene	BPP	Coded Allele	MAF	Unadjusted for ε Alleles		Adjusted for ε Alleles	
						OR [95% CI]	*P* Value	OR [95% CI]	*P* Value
19	rs769449	*APOE*	45410002	A	0.13	1.97 [1.62–2.40]	1.94e−11	1.99 [1.23–3.22]	0.006
19	rs2075650	*TOMM40*	45395619	G	0.15	1.70 [1.42–2.03]	2.14e−08	1.26 [0.91–1.74]	0.15
19	rs4420638	*APOC1*	45422946	G	0.19	1.62 [1.38–1.89]	2.54e−08	1.11 [0.76–1.61]	0.58

ε 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.

detection of a significant association at *P*<1e−06 (corrected for multiple tests in the gene-based analysis) at maximum odds ratio=5 when MAF=0.0001.[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 Center for Public Health Genomics. Dr Anderson received research grants from the NIH-NINDS, American Brain Foundation, and Massachusetts General Hospital 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.

References

1. Devan WJ, Falcone GJ, Anderson CD, Jagiella JM, Schmidt H, Hansen BM, et al; International Stroke Genetics Consortium. Heritability estimates identify a substantial genetic contribution to risk and outcome of intracerebral hemorrhage. *Stroke*. 2013;44:1578–1583. doi: 10.1161/STROKEAHA.111.000089.
2. Woo D, Falcone GJ, Devan WJ, Brown WM, Biffi A, Howard TD, et al; International Stroke Genetics Consortium. Meta-analysis of genome-wide association studies identifies 1q22 as a susceptibility locus for intracerebral hemorrhage. *Am J Hum Genet*. 2014;94:511–521. doi: 10.1016/j.ajhg.2014.02.012.
3. Jeanne M, Labelle-Dumais C, Jorgensen J, Kauffman WB, Mancini GM, Favor J, et al. COL4A2 mutations impair COL4A1 and COL4A2 secretion and cause hemorrhagic stroke. *Am J Hum Genet*. 2012;90:91–101. doi: 10.1016/j.ajhg.2011.11.022.
4. Lee S, Emond MJ, Bamshad MJ, Barnes KC, Rieder MJ, Nickerson DA, et al; NHLBI GO Exome Sequencing Project—ESP Lung Project Team. Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *Am J Hum Genet*. 2012;91:224–237. doi: 10.1016/j.ajhg.2012.06.007.
5. *R: A Language and Environment for Statistical Computing*. Version 3.0.2. Vienna, Austria: R Foundation for Statistical Computing; 2014.
6. Biffi A, Sonni A, Anderson CD, Kissela B, Jagiella JM, Schmidt H, et al; International Stroke Genetics Consortium. Variants at APOE influence risk of deep and lobar intracerebral hemorrhage. *Ann Neurol*. 2010;68:934–943. doi: 10.1002/ana.22134.
7. Donnelly LA, Palmer CN, Whitley AL, Lang CC, Doney AS, Morris AD, et al. Apolipoprotein E genotypes are associated with lipid-lowering responses to statin treatment in diabetes: a Go-DARTS study. *Pharmacogenet Genomics*. 2008;18:279–287. doi: 10.1097/FPC.0b013e3282f60aad.
8. Biffi A, Sonni A, Anderson CD, Kissela B, Jagiella JM, Schmidt H, et al; International Stroke Genetics Consortium. Variants at APOE influence risk of deep and lobar intracerebral hemorrhage. *Ann Neurol*. 2010;68:934–943. doi: 10.1002/ana.22134.
9. Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare-variant association testing for sequencing data with the sequence kernel association test. *Am J Hum Genet*. 2011;89:82–93. doi: 10.1016/j.ajhg.2011.05.029.