Rare and Coding Region Genetic Variants Associated With Risk of Ischemic Stroke
The NHLBI Exome Sequence Project

Paul L. Auer, PhD; Mike Nalls, PhD; James F. Meschia, MD; Bradford B. Worrall, MD; W. T. Longstreth Jr, MD; Sudha Seshadri, MD; Charles Kooperberg, PhD; Kathleen M. Burger, DO; Christopher S. Carlson, PhD; Cara L. Carty, PhD; Wei-Min Chen, PhD; L. Adrienne Cupples, PhD; Anita L. DeStefano, PhD; Myriam Fornage, PhD; John Hardy, PhD; Li F. Hsu, PhD; Rebecca D. Jackson, MD; Gail F. Jarvik, MD, PhD; Daniel S. Kim, BS; Kamakshi Lakshminarayan, MD; Leslie A. Lange, PhD; Ani Manichaikul, PhD; Aaron R. Quinlan, PhD; Andrew B. Singleton, PhD; Timothy A. Thornton, PhD; Deborah A. Nickerson, PhD; Ulrike Peters, PhD; Stephen S. Rich, PhD; for the National Heart, Lung, and Blood Institute Exome Sequencing Project

IMPORTANCE Stroke is the second leading cause of death and the third leading cause of years of life lost. Genetic factors contribute to stroke prevalence, and candidate gene and genome-wide association studies (GWAS) have identified variants associated with ischemic stroke risk. These variants often have small effects without obvious biological significance. Exome sequencing may discover predicted protein-altering variants with a potentially large effect on ischemic stroke risk.

OBJECTIVE To investigate the contribution of rare and common genetic variants to ischemic stroke risk by targeting the protein-coding regions of the human genome.

DESIGN, SETTING, AND PARTICIPANTS The National Heart, Lung, and Blood Institute (NHLBI) Exome Sequencing Project (ESP) analyzed approximately 6000 participants from numerous cohorts of European and African ancestry. For discovery, 365 cases of ischemic stroke (small-vessel and large-vessel subtypes) and 809 European ancestry controls were sequenced; for replication, 47 affected sibpairs concordant for stroke subtype and an African American case-control series were sequenced, with 1672 cases and 4509 European ancestry controls genotyped. The ESP's exome sequencing and genotyping started on January 1, 2010, and continued through June 30, 2012. Analyses were conducted on the full data set between July 12, 2012, and July 13, 2013.

MAIN OUTCOMES AND MEASURES Discovery of new variants or genes contributing to ischemic stroke risk and subtype (primary analysis) and determination of support for protein-coding variants contributing to risk in previously published candidate genes (secondary analysis).

RESULTS We identified 2 novel genes associated with an increased risk of ischemic stroke: a protein-coding variant in PDE4DIP (rs1778155; odds ratio, 2.15; \( P = 2.63 \times 10^{-8} \)) with an intracellular signal transduction mechanism and in ACOT4 (rs35724886; odds ratio, 2.04; \( P = 1.24 \times 10^{-7} \)) with a fatty acid metabolism; confirmation of PDE4DIP was observed in affected sibpair families with large-vessel stroke subtype and in African Americans. Replication of protein-coding variants in candidate genes was observed for 2 previously reported GWAS associations: ZFHX3 (cardioembolic stroke) and ABCA1 (large-vessel stroke).

CONCLUSIONS AND RELEVANCE Exome sequencing discovered 2 novel genes and mechanisms, PDE4DIP and ACOT4, associated with increased risk for ischemic stroke. In addition, ZFHX3 and ABCA1 were discovered to have protein-coding variants associated with ischemic stroke. These results suggest that genetic variation in novel pathways contributes to ischemic stroke risk and serves as a target for prediction, prevention, and therapy.

Published online May 11, 2015.

Copyright 2015 American Medical Association. All rights reserved.
According to the 2010 Global Burden of Disease study, stroke is the second leading cause of death and the third leading cause of years of life lost. Ischemic stroke is the overt symptomatic manifestation of brain infarction, but the burden of cerebrovascular disease is much greater, with many more covert infarctions not resulting in the diagnosis of clinical stroke. More than 7% of asymptomatic adults in the general population have radiographic evidence of brain infarction, with substantially higher rates in the elderly population. Although many effective treatments for stroke exist, novel strategies for stroke prediction, prevention, and therapy need to be found.

Epidemiologic and family studies support an inherited component to stroke risk. Family-based linkage studies, including Cerebral Autosomal Dominant Arteriopathy With Subcortical Infarcts and Leukoencephalopathy and Cerebral Autosomal Recessive Arteriopathy With Subcortical Infarcts and Leukoencephalopathy, identified rare forms of stroke. Ischemic stroke studies include large candidate gene association meta-analyses, mendelian randomization, studies of affected sibpairs, and genome-wide association studies (GWAS). The joint International Stroke Genetics Consortium/Wellcome Trust Case Control Consortium 2 (WTCCC2) effort identified HDAC9 (7p21.1; GenBank NM_058176) and confirmed associations of ischemic stroke with variants in PITX2 (GenBank KJ891816) and ZFHX3 (GenBank NM_006885) (cardioembolic stroke) and in 9p21 (large-vessel/atherosclerotic stroke). The Australian Stroke Genetics Collaborative, the WTCCC2, and the International Stroke Genetics Consortium have identified and replicated the chromosome 6p21.1 large-artery susceptibility locus. Discovering the missing heritability for ischemic stroke could provide critical insights into the cause of the disease, novel pathways, and therapeutic targets.

Rare, protein-coding variation could contribute a burden of risk for ischemic stroke, attributable to multiple variants in a gene rather than any individual variant. Exome sequencing in ischemic stroke has been performed only as a pilot study. Herein, we present results from what we believe to be the first large-scale study of protein-coding region variants in ischemic stroke from the National Heart, Lung, and Blood Institute (NHLBI) Exome Sequencing Project (ESP). We describe analyses used to identify novel stroke susceptibility genes, determine genes and variants that are specific to stroke subtypes, and establish replication of previous GWAS and candidate gene associations.

### Methods

#### Participating Studies and Design

The NHLBI ESP ascertained samples of European American and African American ancestry. Data were obtained on NHLBI cohorts (Atherosclerosis Risk in Communities, Coronary Artery Risk Development in Young Adults, Cardiovascular Health Study, Framingham Heart Study, Jackson Heart Study, the Multi-Ethnic Study of Atherosclerosis, and the Women’s Health Initiative [WHI]) and 2 National Institute of Neurological Disorders and Stroke studies (Siblings With Ischemic Stroke Study [SWISS] and Ischemic Stroke Genetic Study [ISGS])). Characteristics of the participants analyzed are presented in Table 1.

Discovery analysis by exome sequencing focused on small-vessel and large-vessel subtypes in unrelated persons of European ancestry. Replication was conducted in small- and large-vessel subtypes in SWISS and ESP African American individuals, as well as ExomeChip data from European American participants in the WHI that included cardioembolic subtypes.

#### Table 1. Characteristics of Samples Used in the Discovery and Replication Analyses

<table>
<thead>
<tr>
<th>Sample</th>
<th>Cases</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Type of Stroke, No. (%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Small Vessel</td>
<td>Large Vessel</td>
</tr>
<tr>
<td>EA ESP</td>
<td>312/365 (85.5)</td>
<td>62.4 (11.6)</td>
</tr>
<tr>
<td>AA ESP</td>
<td>68/83 (81.9)</td>
<td>57.0 (9.2)</td>
</tr>
<tr>
<td>SWISS</td>
<td>36/94 (38.3)</td>
<td>67.4 (11.5)</td>
</tr>
<tr>
<td>WHI Exome</td>
<td>1672/1672 (100)</td>
<td>68.7 (5.8)</td>
</tr>
</tbody>
</table>

Abbreviations: AA, African ancestry; EA, European ancestry; ESP, Exome Sequencing Project; NA, not available; SWISS, Siblings With Ischemic Stroke Study; WHI, Women’s Health Initiative.
terol levels. Controls for ExomeChip analysis were 4509 WHI participants who self-reported having no history of stroke and myocardial infarction and no family history of stroke.

Ischemic Stroke Classification

Ischemic stroke was defined as a typical clinical syndrome with radiologic confirmation of brain infarction or the absence of an alternative diagnosis for the clinical syndrome. Stroke subtyping was based on the Trial of Org 10172 in Acute Stroke Treatment classification system. Brain imaging by computed tomographic scanning, magnetic resonance imaging, or autopsy confirmation was available to classify all cases. Relevant institutional review boards approved participating studies, and all participants provided written informed consent for participating in the original study as well as for genetic research. There was no financial compensation for use of existing data and samples from the contributing cohorts as part of this study.

Genetic Data

Exome sequencing was performed at the Broad Institute and the University of Washington (eAppendix 2 in the Supplement). ExomeChip genotyping was performed at the Broad Institute (eAppendix 3 in the Supplement). The ESP exome sequencing and ExomeChip genotyping started on January 1, 2010, and continued through June 30, 2012. Analyses were conducted on the full data set between July 12, 2012, and July 13, 2013. A total of 225 239 protein-coding variants were common to both ESP exome sequencing and ExomeChip genotyping, of which 119 963 were polymorphic in the full ESP sample (n = 7355). The self-reported ancestry and genetic ancestry distributions by principal components analysis are shown in eFigure 1 in the Supplement.

Statistical Analysis

Single-variant tests and 2 gene-based tests were used to determine the association with ischemic stroke, conditional on the minor allele frequency (MAF) of individual variants (eAppendix 4 in the Supplement). Variants with MAF of 0.5% or more and at least 100 observations with a nonmissing genotype were analyzed using a single-variant test. For gene-based tests, only missense, splice, or nonsense variants were considered. Variants with MAF of 5% or less were analyzed with the sequence kernel association test (SKAT), or variants with MAF of 1% or less analyzed with the combined multivariate and collapsing method, using variants with MAF of less than 1% (CMC/T1 test). For SKAT and CMC/T1, only genes with cumulative MAF of 0.5% or more were assessed. Analyses were conducted separately for exome sequence (combining data from all cohorts and ISGS) and ExomeChip (WHI) data; a meta-analysis was conducted using seqMeta software (http://cran.r-project.org/web/packages/seqMeta2). Covariate adjustment included participant’s age at baseline and sex, as appropriate. For autosomal variants, a log-additive genetic model was used; for sex chromosome variants, a dominance model was used.

Ischemic stroke and subtype analyses used 2037 cases and 5318 controls with ESP exome sequencing or ExomeChip data. Association of ischemic stroke with any single variant used \( P < 4.17 \times 10^{-7} \) (0.05/119 963) as the threshold for exomewide significance. Restriction to variants with MAF of 0.005 or more and at least 100 observations with a nonmissing genotype yielded 25 467 variants, with a significance threshold at \( P < 1.96 \times 10^{-6} \). Gene-level associations used a variety of statistical tests and MAF cutoffs; CMC/T1 is most powerful when all variants in a gene influence risk in the same direction; SKAT is more powerful when the variants in a gene influence risk in opposite directions. For CMC/T1, genes were included for individual variants with MAF of 0.01 or less and cumulative MAF of 0.005 or more, resulting in 6173 informative genes with a threshold of \( P < 8.10 \times 10^{-6} \). For SKAT, there were 8361 genes with MAF of 0.05 or less and cumulative MAF of 0.005 or more, with a threshold of \( P < 5.98 \times 10^{-6} \). In the discovery phase, these analyses were conducted for all ischemic stroke, small-vessel subtype, and large-vessel subtype. Affected sibpairs from SWISS were analyzed using identity-by-descent methods (eAppendix 4 in the Supplement). Regions of excess identity by descent were examined, focusing on missense, nonsense, stop-gain/loss functional variants. Single-variant and aggregate identity-by-descent probabilities of each gene across all available sibpairs were calculated to generate logarithm-of-the-odds scores for excess identity-by-descent sharing, adjusting for local recombination rates based on European ancestry estimates (HapMap phase 3), repeated for sibpairs concordant for small-vessel and large-vessel subtypes. The distribution of scores from the identity-by-descent analyses is shown in eFigure 2 in the Supplement.

Targeted secondary analysis of 59 candidate genes previously associated with ischemic stroke, stroke risk factors, and mendelian syndromes yielded 134 variants. Most of these variants are not located in protein-coding regions; however, proxies were identified with strong linkage disequilibrium with ESP-defined protein-coding variants. For single-variant tests, significance was defined as \( P < 1.86 \times 10^{-4} \). For gene-based tests, 33 genes contained polymorphic variants for analysis and significance was defined as \( P < 7.58 \times 10^{-4} \) since 2 gene-based tests (CMC/T1 and SKAT) were used.

Results

Exome Variants Associated With Ischemic Stroke

Single-variant tests identified 2 protein-coding variants significantly associated with ischemic stroke (Table 2). A common (MAF, 0.316) missense variant, rs1778155, in PDE4DIP (phosphodiesterase 4D-interacting protein; 1q21.1; GenBank NM_014644) was associated with ischemic stroke (odds ratio [OR], 2.15; 95% CI, 1.65 to 2.78). An infrequent (MAF, 0.017) missense variant, rs35724886, in ACOT4 (acyl-coenzyme A thioesterase 4; 14q24.3; GenBank NM_152331) was associated with ischemic stroke (OR, 2.04; 95% CI, 1.29 to 3.23). Although not reaching exomewide significance, a common (MAF, 0.379) variant, rs5001076, in primase, DNA, polypeptide 2 (58 kilodaltons [kDa]); 6p11.2; GenBank NM_000947) showed evidence of increased risk with ischemic stroke (OR, 2.11; 95% CI, 1.39 to 3.16) (Table 2).
Rare and infrequent protein-coding variants in *PDE4DIP* from exome sequencing in 365 cases and 809 controls and their estimated effect on ischemic stroke risk are shown in the Figure. A large complex gene (240.030 kilobases [kb]; 2346 amino acids, 265 kDa), *PDE4DIP* has 86 distinct introns, 64 different messenger RNAs, 49 alternatively spliced variants, and at least 10 transcription factor binding sites in its promoter. In contrast, *ACOT4* is a small gene (4.791 kb; 421 amino acids, 46 kDa) with 2 distinct introns and 1 spliced messenger RNA. Both genes are expressed in the brain and in other tissues.

The results of no single-variant or gene-based test attained statistical significance for the small-vessel (lacunar) stroke subtype. *PDE4DIP* rs1778155 had direction and size of effect on risk consistent in the all-stroke and small-vessel stroke categories (OR, 2.03; meta-analysis $P = 7.96 \times 10^{-6}$). *ACOT4* rs35724886 also had the same direction and similar size of effect on risk (OR, 1.95; meta-analysis $P = 1.36 \times 10^{-7}$).

### Table 2. Association Results for Coding Region Variants and Risk of Ischemic Stroke With a Meta-analysis $P < 5 \times 10^{-6}$ Displayed

<table>
<thead>
<tr>
<th>Gene</th>
<th>Location</th>
<th>Gene Size, kb</th>
<th>Associated SNV</th>
<th>MAF</th>
<th>OR</th>
<th>Replication</th>
<th>Meta-analysis</th>
<th>No. of SNVs in SKAT</th>
<th>SKAT Gene-Level P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDE4DIP</td>
<td>1q21.1</td>
<td>240.030</td>
<td>rs1778155</td>
<td>0.316</td>
<td>2.147</td>
<td>NA</td>
<td>2.147</td>
<td>2.63 $\times 10^{-6}$</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>42</td>
<td>.97</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>24</td>
<td>.43</td>
</tr>
<tr>
<td>ACOT4</td>
<td>14q24.3</td>
<td>4.791</td>
<td>rs35724886</td>
<td>0.017</td>
<td>2.044</td>
<td>NA</td>
<td>2.044</td>
<td>1.24 $\times 10^{-7}$</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1.36 $\times 10^{-7}$</td>
</tr>
<tr>
<td>PRIM2</td>
<td>6p11.2</td>
<td>333.773</td>
<td>rs5001076</td>
<td>0.379</td>
<td>2.105</td>
<td>NA</td>
<td>2.105</td>
<td>2.96 $\times 10^{-6}$</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td>7.2 $\times 10^{-5}$</td>
</tr>
</tbody>
</table>

### Table 3. Association Results for Coding Region Variants and Risk of Small-Vessel Ischemic Stroke With Meta-analysis $P < 5 \times 10^{-4}$ Displayed

<table>
<thead>
<tr>
<th>Gene</th>
<th>Location</th>
<th>Gene Size, kb</th>
<th>SNV</th>
<th>MAF</th>
<th>OR ESP</th>
<th>ESP</th>
<th>Exome-Chip</th>
<th>Meta-analysis</th>
<th>P Value</th>
<th>ESP</th>
<th>Exome-Chip</th>
<th>Meta-analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDE4DIP</td>
<td>1q21.1</td>
<td>240.030</td>
<td>rs1778155</td>
<td>0.308</td>
<td>2.029</td>
<td>NA</td>
<td>2.029</td>
<td>7.96 $\times 10^{-6}$</td>
<td>NA</td>
<td>7.96 $\times 10^{-6}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACOT4</td>
<td>14q24.3</td>
<td>4.791</td>
<td>rs35724886</td>
<td>0.018</td>
<td>1.952</td>
<td>NA</td>
<td>1.952</td>
<td>2.43 $\times 10^{-5}$</td>
<td>NA</td>
<td>2.43 $\times 10^{-5}$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Gene-Based Test Used (SKAT)

<table>
<thead>
<tr>
<th>Gene</th>
<th>Location</th>
<th>Gene Size, kb</th>
<th>SNV</th>
<th>MAF</th>
<th>Cumulative, P Value</th>
<th>ESP</th>
<th>Exome-Chip</th>
<th>Meta-analysis</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CEP164</td>
<td>11q23.3</td>
<td>98.712</td>
<td>n = 24</td>
<td>0.010</td>
<td>0.010</td>
<td>0.010</td>
<td>0.010</td>
<td>0.010</td>
<td>0.997</td>
</tr>
</tbody>
</table>

**Abbreviations:** kb, kilobase; MAF, minor allele frequency; NA, not applicable; OR, odds ratio; SKAT, sequence kernel association test; SNV, single-nucleotide variant.
Table 4. Association Results for Coding Region Variants and Risk of Large-Vessel Ischemic Stroke With Meta-analysis P<5×10⁻⁴ Displayed

<table>
<thead>
<tr>
<th>Gene</th>
<th>Location</th>
<th>Gene Size, kb</th>
<th>SNV</th>
<th>Test Results</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>MAF</td>
<td>OR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPSB3</td>
<td>16p13.3</td>
<td>16.989</td>
<td>Best, rs147735377</td>
<td>0.006</td>
<td>NA</td>
</tr>
<tr>
<td>PDE4DIP</td>
<td>1q21.1</td>
<td>240.030</td>
<td>Best, rs1778155</td>
<td>0.302</td>
<td>2.40</td>
</tr>
<tr>
<td>DNAH14</td>
<td>1q42.12</td>
<td>503.033</td>
<td>Best, rs41267347</td>
<td>0.080</td>
<td>2.33</td>
</tr>
<tr>
<td>MPDZ</td>
<td>9p23</td>
<td>173.887</td>
<td>Best, rs188840960</td>
<td>0.009</td>
<td>2.51</td>
</tr>
<tr>
<td>ZNF778</td>
<td>16q24.3</td>
<td>11.855</td>
<td>Best, rs117690401</td>
<td>0.040</td>
<td>4.380</td>
</tr>
<tr>
<td>SLC22A5</td>
<td>5q31.1</td>
<td>25.906</td>
<td>Best, rs28383481</td>
<td>0.005</td>
<td>0.9289</td>
</tr>
<tr>
<td>PGAP1</td>
<td>2q33.1</td>
<td>94.793</td>
<td>Best, rs62185645</td>
<td>0.009</td>
<td>23.482</td>
</tr>
<tr>
<td>KIF21A</td>
<td>12q12</td>
<td>150.163</td>
<td>Best, rs78616703</td>
<td>0.005</td>
<td>2.44</td>
</tr>
<tr>
<td>MX1</td>
<td>21q22.3</td>
<td>38.911</td>
<td>Best, rs140022520</td>
<td>0.005</td>
<td>51.953</td>
</tr>
<tr>
<td>ACOT4</td>
<td>14q24.3</td>
<td>4.791</td>
<td>Best, rs35724886</td>
<td>0.015</td>
<td>2.384</td>
</tr>
</tbody>
</table>

Abbreviations: ESP, Exome Sequencing Project; kb, kilobase; MAF, minor allele frequency; NA, not applicable; OR, odds ratio; SNV, single-nucleotide variant; SKAT, sequence kernel association test.

P = 2.43×10⁻⁵. One novel gene, CEPT64 (centrosomal protein 164 kDa; 11q23.3; GenBank NM_014956) with 24 rare variants (cumulative MAF, 0.01) approached statistical significance (SKAT meta-analysis P = 7.93×10⁻⁶).

Within large-vessel (atherosclerotic) stroke, no single-variant or gene-based test attained statistical significance (Table 4). PDE4DIP rs1778155 had a strong effect and consistent direction (OR, 2.40; P = 2.18×10⁻⁵). Single-variant tests did not support ACOT4 rs35724886, but gene-based tests of many infrequent or rare variants provided suggestive evidence for association (eg, SPSB3 [GenBank KJ899913], SLC22A5 [GenBank KJ897580], PGAP1 [GenBank NM_024989], KIF21A [GenBank NM_001173464], and MX1 [GenBank NM_001144925]) with large-vessel subtype.

Cases of cardioembolic ischemic stroke were not selected for ESP exome sequencing; these participants were sequenced only if they were selected for other (nonstroke) phenotypes or were included later for ExomeChip genotyping. No statistically significant associations with cardioembolic stroke were found by either single-variant or gene-based tests. The most significant single-variant association was with PIP5K2 rs35671301 (GenBank NM_001276277) (diphosphoinositol pentakisphosphate kinase 2; 5q21.1; OR, 3.59; MAF, 0.011; meta-analysis P = 1.01×10⁻⁵). The SKAT test also provided support for PIP5K2 on risk in the cardioembolic stroke (5 variants with cumulative MAF, 0.104; meta-analysis P = 1.51×10⁻⁵).

Replication of Exome Variants Associated With Ischemic Stroke

Single-nucleotide polymorphisms in PDE4DIP and ACOT4 were analyzed in an independent set of affected sibpairs from SWISS and unrelated ESP African American participants. The PDE4DIP rs1778155 variant had significantly increased allele sharing identity by descent (P = 1.16×10⁻⁴) and was observed in 16 of 19 large-vessel concordant pairs with a logarithm of the odds of 6.07 (OR, 2.41; MAF, 0.302; P = 2.18×10⁻⁵) and a false discovery rate of P = 6.29×10⁻³. ACOT4 rs35724886 was not polymorphic in either the SWISS or WHI ExomeChip data. A significant association with ischemic stroke was observed for PDE4DIP rs1778155 in the ESP African American samples (MAF, 0.191; OR, 2.34, P = 7.50×10⁻⁴). This variant failed genotype quality control metrics on the ExomeChip and was not analyzed. In African Americans, no significant association was detected for...
ACOT4 rs35724886 with ischemic stroke although the effect size was similar (OR, 1.86; MAF, 0.180; $P = .31$).

**Analysis of Previously Identified Candidate Genes and Variants**

We identified 59 published candidate genes to be analyzed with the single-variant and gene-based tests. A single variant in ABCA1 (adenosine triphosphate–binding cassette, subfamily A [ABCA1], member 1; 9q31.1; GenBank NM_005502) was associated with all ischemic stroke (meta-analysis $P = 1.94 \times 10^{-4}$) and in both subtypes (small-vessel meta-analysis $P = 0.008$; large-vessel meta-analysis $P = 0.02$). Variants in ZFHX3 (zinc finger homeobox 3; 16q22.3), a candidate gene for atrial fibrillation, were nominally associated with all ischemic stroke using the CMC/T1 test ($P = .001$), primarily in the cardioembolic subtype ($P = .04$). Results for all candidate genes are presented in eTables 1 through 4 in the Supplement.

**Discussion**

We report the results of ESP exome sequencing and ExomeChip genotyping of 2037 cases of ischemic stroke and 5318 controls, with replication in an African American population and a collection of affected sibpairs. We present evidence supporting the role of protein-coding variants in 2 novel genes, PDE4DIP and ACOT4, which are associated with an increased risk of ischemic stroke. We evaluated protein-coding variants in previously reported genes associated with ischemic stroke, and support the role of ABCA1 in large-vessel stroke and ZFHX3 in cardioembolic stroke.

A primary interacting partner of PDE4DIP is PDE4D (phosphodiesterase 4D, cAMP-specific; GenBank NM_001104631), originally identified as linked to ischemic stroke in Icelandic families. The PDE4D variants have been extensively studied for association with ischemic stroke risk in diverse populations, with equivocal results. A meta-analysis of ischemic stroke was conducted on 7 studies, using approximately 12,000 cases and 15,000 controls that included 6 single-nucleotide polymorphisms in PDE4D. In the meta-analysis, PDE4D rs702553 had a significant association with ischemic stroke with the PDE4D result, coupled with PDE4DIP identified in the present study, supports a role of this pathway in ischemic stroke.

The ACOT4 gene encodes an enzyme that catalyzes the hydrolysis of acyl-coenzyme A (CoA) to the free fatty acid and coenzyme A, regulating intracellular levels of acyl-CoA and fatty acids. Selected fatty acids are associated with adverse outcomes including dyslipidemia, inflammation, myocardial infarction, and other cardiovascular mortality. Human ACOT4 hydrolyzes succinyl-CoA, glutaryl-CoA, and long-chain acyl-CoA reactions catalyzed by multiple enzymes in other species. Expression of ACOT4 is regulated by peroxisome proliferator-activated receptor-α (PPARα), a nuclear receptor that has been implicated in vascular and cardiac disease and represents a therapeutic target in diabetes mellitus. ACOT4 links risk for ischemic stroke with fatty acid metabolism.

This study, which we believe to be the largest on exome sequencing in stroke, has limitations. The replication of PDE4DIP rs1778155 association with ischemic stroke succeeded in both affected sibpair families and a small case-control cohort of African ancestry; however, it was not replicated in ExomeChip analyses. Furthermore, the ACOT4 rs35724886 infrequent variant was not informative in the replication samples and, therefore, was not replicated. Because these are protein-coding variants, their coverage in GWAS reports (eg, METASTROKE) is often incomplete. PDE4DIP rs1778155 has no surrogate within $r^2$ value greater than 0.4 in the 1000 Genomes Project data; ACOT4 rs35724886 has no surrogate within $r^2$ value greater than 0.8 in the ACOT2/ACOT4/ACOT6 cluster. In the joint ImmunoChip/WTCCC2/METASTROKE data, the PDE4DIP rs1778155 variant is 20.5 Mb proximal to the SELP (selectin P [granule membrane protein 140kDa, antigen CD62; 1q24.2]; GenBank NM_003005) rs3917792 variant (intronic, associated with all ischemic stroke: OR, 1.14; $P = 1.76 \times 10^{-7}$). However, ACOT4 rs35724886 is only 1.4 Mb from the METASTROKE-identified RGS6 (regulator of G-protein signaling 6; 14q24.2); GenBank NM_001204416) rs2238238 variant (intronic, associated with all ischemic stroke: OR, 1.18; $P = 1.65 \times 10^{-7}$). Thus, METASTROKE provides indirect support for the ACOT4 contribution to ischemic stroke. Several candidate genes and loci were not analyzed, since few protein-coding variants were identified through ESP exome sequencing or included on the ExomeChip (eg, HDAC9, PDE4D). The 9p21.3 region, previously associated with large-vessel ischemic stroke, contains few protein-coding genes and, therefore, few coding-region variants. Only 3 variants in CDKN2A (cyclin-dependent kinase inhibitor 2A; GenBank NG_007485) and 1 variant in CDKN2B-AS1 (CDKN2B-antisense RNA 1; GenBank HG975381) were analyzed, with no variant associated with overall ischemic stroke or stroke subtype. Finally, our discovery sample focused on small- and large-vessel ischemic stroke subtypes to increase statistical power to detect novel genetic loci and reduce heterogeneity. Subsequent epidemiologically robust sampling approaches, consistent with the underlying distribution of subtypes in the population, should be conducted. Such studies would enable estimation of risk attributable to these genes (eg, PDE4DIP and ACOT4) and pathways to small- and large-vessel stroke and the effect of these variants in individual risk profiles for ischemic stroke.

The NHLBI ESP is the first large-scale exome sequencing project to focus on complex human phenotypes rather than mendelian disorders. Although data from both European Americans and African Americans were sequenced in the ESP, only European Americans had sufficiently large numbers of ischemic stroke to permit robust genetic analyses. The motivation for exploring the exome was that the protein-coding region variants would have greater likelihood of functional impact and, therefore, larger effects on disease risk, although examples of this remain few and require large sample sizes, similar to those of GWAS. Nevertheless, exome sequencing has, as exhibited in the present study, provided novel gene targets and biological pathways for examination on their role in ischemic stroke.
Conclusions

Our study has led to the identification of 2 novel genes (PDE4DIP and ACOT4) and the replication of 2 previously reported candidate genes (ABCA1 and ZFXH3) as containing coding region variants associated with ischemic stroke. These new results suggest that pathways, involving cell migration and growth (PDE4DIP) and long-chain fatty acid metabolism (ACOT4), could provide insights into the cause of ischemic stroke and as targets for pharmacologic interventions and therapies.

Funding/Support: Funding for GO ESP was provided by NHLBI grants RC2 HL-103010 (HeartGO), RC2 HL-102923 (LungGO), and RC2 HL-102924 (Women’s Health Initiative Sequencing Project). The exome sequencing was performed through NHLBI grants RC2 HL-102925 (BroadGO) and RC2 HL-102926 (SeattleGO). HeartGO components and their support include Atherosclerosis Risk in Communities (NHLBI contracts N01-HC-55015, N01-HC-55016, N01-HC-55017, N01-HC-55018, N01-HC-55019, N01-HC-55020, and N01-HC-55021), Cardiovascular Health Study (NHLBI contracts HHSN268201200003EC, HHSN2682008000007C, N01HC55222, N01HC85307, N01HC85308, N01HC85301, N01HC85302, N01HC85303, and N01HC85306), and NHLBI grants U01HL082950, RO1HL076392, RO1HL105765, RO1HL106162, and RO1HL103934, with additional contribution from the National Institute of Neurological Disorders and Stroke. Additional support was provided through R21AG023629 from the National Institute on Aging. A full list of principal Cardiovascular Health Study investigators and institutions can be found at CHS-NHLBI.org; Coronary Artery Risk Development in Young Adults (NHLBI contracts N01-HC-95095, N01-HC-48047, N01-HC-48048, N01-HC-48049, and N01-HC-48050); Framingham Heart Study (NHLBI contract N01-HC-25195 and grants HS75790, AG08122, and AG033193); Jackson Heart Study (NHLBI contracts N01 HC-95170, N01 HC-95171, and N01 HC-95172); Multi-Ethnic Study of Atherosclerosis (NHLBI contracts N01-HC-95159 through N01-HC-95169 and grant 024156). The Siblings with Ischemic Stroke Study (grant R01-NS39987) and the Ischemic Stroke Genetics Study (grant R01-NS42773) contributed phenotypic data and DNA samples. Dr Nalls’ participation in this research was supported, in part, by the Intramural Research Program of the National Institute on Aging (grant Z01-AG00954-7). Portions of this study used the high-performance computational capabilities of the Biowulf Linux cluster at the National Institutes of Health (http://biowulf.nih.gov). The Women’s Health Initiative program is funded by the NHLBI through contracts HHSN2682012000046C, HHSN268201000001C, HHSN268201200002C, HHSN268201000003C, HHSN268201200005C, HHSN268201200006C, HHSN268201200007C, and HHSN268201200008C. The Lung GO components and their support include Cystic Fibrosis Foundation (Cystic Fibrosis Foundation grants GBS00170X, KN04160X, OBSERV0401X, and RRP026X), and NIH grants R01-HL068890, R22-HL095396, U1L-RR025014, and S000-HGO04316); Chronic Obstructive Pulmonary Disease (COPDGene; NIH grants U01-HL089897 and U01-HL089856), and the COPD Foundation through contributions made to an industry advisory board comprising AstraZeneca, Boehringer Ingelheim, Novartis, Pfizer, and Sunovion; Acute Lung Injury (NIH grant RC2-HL017779); Lung Health Study (NIH grants R2C-HL065853, R01-HG004738, and HR-46002); Pulmonary Arterial Hypertension (NIH grants P50-HL084946, K23-AR52742, and F32-HL083714); and Asthma (NIH grants RC2-HL106561, HL077716, HL-65917, HL-76285, and MO1-RR07222).

Role of the Funder/Sponsor: The funding organization (NHLBI) had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Group Information: NHLBI Exome Sequencing Project members are listed in eAppendix 1 in the Supplement.

Additional Contributions: The support of the research institutions, study investigators, field staff, and study participants assisting in this creation for biomedical research.

REFERENCES


