Identification of Novel Loci for Alzheimer Disease and Replication of CLU, PICALM, and BIN1 in Caribbean Hispanic Individuals

Joseph H. Lee, DrPH; Rong Cheng, PhD; Sandra Barral, PhD; Christiane Reitz, MD, PhD; Martin Medrano, MD; Rafael Lanituga, MD; Ivonne Z. Jiménez-Velázquez, MD; Ekaterina Rogaeva, PhD; Peter H. St. George-Hyslop, MD; Richard Mayeux, MD, MSc

Objectives: To identify novel loci for late-onset Alzheimer disease (LOAD) in Caribbean Hispanic individuals and to replicate the findings in a publicly available data set from the National Institute on Aging Late-Onset Alzheimer’s Disease Family Study.

Design: Nested case-control genome-wide association study.


Participants: Five hundred forty-nine affected and 544 unaffected individuals of Caribbean Hispanic ancestry.

Intervention: The Illumina HumanHap 650Y chip for genotyping.

Main Outcome Measure: Clinical diagnosis or pathologically confirmed diagnosis of LOAD.

Results: The strongest support for allelic association was for rs9945493 on 18q23 ($P = 1.7 \times 10^{-7}$), but 22 additional single-nucleotide polymorphisms (SNPs) had a $P$ value less than $9 \times 10^{-6}$ under 3 different analyses: unadjusted and stratified by the presence or absence of the APOE ε4 allele. Of these SNPs, 5 SNPs (rs4669573 and rs10197851 on 2p25.1; rs11711889 on 3q25.2; rs1177750 on 7p21.1; and rs7908652 on 10q23.1) were associated with LOAD in an independent cohort from the National Institute on Aging Late-Onset Alzheimer’s Disease Family Study. We also replicated genetic associations for CLU, PICALM, and BIN1.

Conclusions: Our genome-wide search of Caribbean Hispanic individuals identified several novel genetic variants associated with LOAD and replicated these associations in a white cohort. We also replicated associations in CLU, PICALM, and BIN1 in the Caribbean Hispanic cohort.


Numerous genome-wide association studies (GWAS) have been published for late-onset Alzheimer disease (LOAD).1-13 Aside from APOE, additional candidate susceptibility genes identified using GWAS methods for LOAD have included GAB2, GALP, 14q32.13, LOC651924, PGBD1, TNK1, CR1, CLU, PICALM, and BIN1.14,15 In addition, variants in SORL1 identified by Rogaeva et al16 have been replicated in several independent cohorts and were significantly associated with LOAD in a meta-analysis.17 Difficulties inherent to the genetics of complex diseases (eg, etiologic heterogeneity, gene × environment and gene × gene interactions, and methylation) remain with these studies, and much work needs to be done. For example, the strength of association, or effect size, as measured by odds ratios (ORs) varies widely across studies and is generally small. Yet, these GWAS have identified a number of candidate genes that need to be replicated and their functional roles determined. Despite the increasing number of identified susceptibility genetic variants, a relatively large proportion of genetic variance remains unexplained.18 This has much to do with both the complexity of the genetics and inadequacy of heritability as a measure of genetic contribution. Similar phenomena have been observed in other common, complex genetic diseases and invoked a term, genetic dark matter, in GWAS.19,20

In the current study, we report the results of a GWAS in unrelated patients with LOAD and controls of Caribbean Hispanic ancestry. This population was selected because the prevalence and incidence rate of LOAD is higher than in white, non-Hispanic individuals living in the same community11 and because we had previously identified numerous large families

Author Affiliations are listed at the end of this article.
multiply affected by LOAD. We first examined unrelated cases and controls in the Caribbean Hispanic individuals and then replicated the associations using the publicly available GWAS data from the National Institute on Aging Late-Onset Alzheimer’s Disease (NIA-LOAD) Family Study (E. M. Wijsman, PhD, N. Pankratz, PhD, Y. Choi, PhD, J. H. Rothstein, MS, K. Faber, MS, R.C., J.H.L., T. D. Bird, MD, D. A. Bennett, MD, R. Diaz-Arrastia, MD, A. M. Goate, DPhil, M. Farlow, MD, B. Gheiti, MD, R. A. Sweet, MD, T. M. Foroud, PhD, and R.P.M.; for the NIA-LOAD/NCRAD Family Study Group. “Genome-wide Association of Familial Late-Onset Alzheimer’s Disease Replicates BIN1 and CLU and Nominates CUGBP2 in Interaction with APOE,” unpublished data). This approach allowed us to further assess the role of genetic admixture in the Caribbean Hispanic population. To our knowledge, this is the only GWAS of Alzheimer disease that focuses exclusively on a Caribbean Hispanic population.

**METHODS**

**SAMPLES OF CARIBBEAN HISPANIC INDIVIDUALS**

We studied 1093 unrelated Caribbean Hispanic individuals comprising 549 cases and 544 controls (Table 1). These participants were selected from the Washington Heights–Inwood Columbia Aging Project (WHICAP) study and the Estudio Familiar de Influencia Genetica de Alzheimer (EFIGA) study. The WHICAP study is a population-based epidemiologic study of randomly selected elderly individuals residing in northern Manhattan, New York, comprising 3 ethnic groups: non-Hispanic white, Caribbean Hispanic, and African American. For the current study, we restricted the study inclusion to individuals who were self-reported Hispanic of Caribbean origin and did not include non-Hispanic white or African American individuals. In addition, we selected 1 affected individual from each family participating in the EFIGA study of Caribbean Hispanic families with LOAD. Both studies followed the same clinical diagnostic methods.

The participants originated from the Dominican Republic and Puerto Rico. Approximately 60.3% of the affected individuals were participants in the WHICAP epidemiologic study, and the remaining 39.7% of the participants were from the EFIGA study. All unaffected individuals were participants in the WHICAP epidemiologic study. For the familial cases, we selected 1 proband from each family to create a cohort of unrelated individuals. We selected persons with definite or probable LOAD over those with possible LOAD to limit the effects of comorbidity.

**CLINICAL ASSESSMENTS**

Data were available from medical, neurological, and neuropsychological evaluations collected from 1999 through 2007. The standardized neuropsychological test battery covered multiple domains and included the Mini-Mental State Examination, the Boston Naming Test, the Controlled Word Association Test from the Boston Diagnostic Aphasia Evaluation, the Wechsler Adult Intelligence Scale–Revised similarities subtest, the Mattis Dementia Rating Scale, the Rosen Drawing Test, the Benton Visual Retention Test, the multiple-choice version of the Benton Visual Retention Test, and the Selective Reminding Test.

**DIAGNOSIS OF DEMENTIA**

The diagnosis of dementia was established on the basis of all available information gathered from the initial and follow-up assessments and medical records. The diagnosis of LOAD was based on the National Institute of Neurological Disorders and Stroke–Alzheimer’s Disease and Related Disorders Association criteria.

**GENOTYPING**

Single-nucleotide polymorphisms (SNPs) were genotyped at the Illumina Genotyping Service Center, San Diego, California, using Illumina HumanHap 650Y chips. From the 650Y chips, 658 610 SNP markers were originally genotyped. Quality control measures for SNP genotype were performed using PLINK (http://pngu.mgh.harvard.edu/~purcell/plink/). We excluded SNPs with the following characteristics: missing genotype rate more than 20%; minimum allele frequency less than 1%; Hardy-Weinberg equilibrium test at a P value less than .0001 in controls. Although the 650Y chip includes additional SNPs for Yoruban individuals, we initially used less stringent criteria for quality control than others because the Illumina SNP chips are optimized for white populations. Furthermore, we wanted to reduce the likelihood of false-negative results. To limit the possibility that positive signals were caused by SNPs with poor calling rate, we lowered the threshold for the missing genotype rate to 5%. This screen reduced the total number of analyzed SNPs by 0.26%. None of the SNPs of main interest (ie, P value < 9 × 10<sup>−8</sup> shown in Table 2) had low genotype rates. Following all quality control measures, we analyzed 627 380 autosomal SNPs.

**POPULATION STRATIFICATION**

We applied 2 methods to estimate ancestry proportion in each subject, and thus population stratification, in this case-control data set: STRUCTURE version 2.2 and identity-by-state–
Table 2. Candidate SNPs From the Caribbean Hispanic GWAS and Replication in the NIA-LOAD GWAS

<table>
<thead>
<tr>
<th>Chr</th>
<th>SNP</th>
<th>bp</th>
<th>Cyto</th>
<th>All</th>
<th>(\epsilon^4)</th>
<th>Unrelated</th>
<th>Family-Based Association</th>
<th>Flanking SNP</th>
<th>Candidate Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 rs7525939</td>
<td>224 822 594</td>
<td>1q42.1</td>
<td>5.29 (\times 10^{-4})</td>
<td>0.00462</td>
<td>0.01300</td>
<td>0.48650</td>
<td>0.67635</td>
<td>0.16364</td>
</tr>
<tr>
<td></td>
<td>2 rs4669573^a</td>
<td>10 396 387</td>
<td>2p25.1</td>
<td>5.26 (\times 10^{-4})</td>
<td>3.67 (\times 10^{-6})</td>
<td>0.02200</td>
<td>0.09629</td>
<td>0.36830</td>
<td>0.25040</td>
</tr>
<tr>
<td>2</td>
<td>rs10197851</td>
<td>10 402 860</td>
<td>2p25.1</td>
<td>0.00102</td>
<td>7.13 (\times 10^{-6})</td>
<td>0.54250</td>
<td>0.20220</td>
<td>0.71710</td>
<td>0.10894</td>
</tr>
<tr>
<td>3</td>
<td>rs1402752</td>
<td>44 662 304</td>
<td>3p21.31</td>
<td>0.00663</td>
<td>64.2500</td>
<td>8.69 (\times 10^{-6})</td>
<td>0.80600</td>
<td>0.88493</td>
<td>0.81175</td>
</tr>
<tr>
<td>3</td>
<td>rs11711889^a</td>
<td>154 853 356</td>
<td>3q25.2</td>
<td>6.95 (\times 10^{-4})</td>
<td>0.00175</td>
<td>0.02247</td>
<td>0.03814</td>
<td>0.94935</td>
<td>NA</td>
</tr>
<tr>
<td>5</td>
<td>rs919289</td>
<td>120 593 422</td>
<td>5q33.1</td>
<td>0.00468</td>
<td>3.20 (\times 10^{-4})</td>
<td>0.47380</td>
<td>0.39070</td>
<td>0.34351</td>
<td>0.95958</td>
</tr>
<tr>
<td>5</td>
<td>rs4895298</td>
<td>120 597 007</td>
<td>5q33.1</td>
<td>0.00477</td>
<td>2.00 (\times 10^{-6})</td>
<td>0.65480</td>
<td>0.63350</td>
<td>0.22174</td>
<td>0.14639</td>
</tr>
<tr>
<td>5</td>
<td>rs2973413</td>
<td>120 597 210</td>
<td>5q33.1</td>
<td>0.00571</td>
<td>5.53 (\times 10^{-4})</td>
<td>0.59630</td>
<td>0.53150</td>
<td>0.21149</td>
<td>0.15152</td>
</tr>
<tr>
<td>7</td>
<td>rs10271466</td>
<td>10 090 002</td>
<td>7p13.31</td>
<td>0.00010</td>
<td>0.00005</td>
<td>0.02260</td>
<td>0.00915</td>
<td>0.03920</td>
<td>NA</td>
</tr>
<tr>
<td>7</td>
<td>rs1117750^a</td>
<td>1 485 943</td>
<td>7p21.1</td>
<td>8.02 (\times 10^{-4})</td>
<td>0.02861</td>
<td>0.01305</td>
<td>0.95270</td>
<td>0.45630</td>
<td>0.26433</td>
</tr>
<tr>
<td>8</td>
<td>rs11786902</td>
<td>4 388 532</td>
<td>8q22</td>
<td>0.00134</td>
<td>8.67 (\times 10^{-4})</td>
<td>0.97670</td>
<td>0.00000</td>
<td>0.43359</td>
<td>0.91219</td>
</tr>
<tr>
<td>8</td>
<td>rs6477258</td>
<td>8 085 638</td>
<td>9q24.1</td>
<td>7.19 (\times 10^{-4})</td>
<td>0.002497</td>
<td>0.001142</td>
<td>0.00000</td>
<td>0.00000</td>
<td>NA</td>
</tr>
<tr>
<td>9</td>
<td>rs10758939</td>
<td>8 115 410</td>
<td>9q24.1</td>
<td>4.75 (\times 10^{-4})</td>
<td>0.00478</td>
<td>0.004208</td>
<td>0.68920</td>
<td>0.25155</td>
<td>0.36560</td>
</tr>
<tr>
<td>10</td>
<td>rs7908652^a</td>
<td>85 781 903</td>
<td>10q23.1</td>
<td>3.59 (\times 10^{-4})</td>
<td>0.02532</td>
<td>0.00451</td>
<td>0.02650</td>
<td>0.13064</td>
<td>0.3457</td>
</tr>
<tr>
<td>11</td>
<td>rs978769</td>
<td>110 393 694</td>
<td>11q23.1</td>
<td>2.90 (\times 10^{-4})</td>
<td>0.003084</td>
<td>0.000287</td>
<td>0.67400</td>
<td>0.05720</td>
<td>0.25093</td>
</tr>
<tr>
<td>11</td>
<td>rs978770</td>
<td>110 393 802</td>
<td>11q23.1</td>
<td>1.49 (\times 10^{-4})</td>
<td>0.002075</td>
<td>0.000223</td>
<td>0.68270</td>
<td>0.05678</td>
<td>0.25242</td>
</tr>
<tr>
<td>11</td>
<td>rs11213703</td>
<td>110 395 150</td>
<td>11q23.1</td>
<td>2.35 (\times 10^{-4})</td>
<td>0.002643</td>
<td>0.000287</td>
<td>0.63300</td>
<td>0.06128</td>
<td>0.30704</td>
</tr>
<tr>
<td>13</td>
<td>rs11617026</td>
<td>100 843 480</td>
<td>13q33.1</td>
<td>0.00912</td>
<td>0.380700</td>
<td>1.70 (\times 10^{-6})</td>
<td>0.44530</td>
<td>0.81524</td>
<td>0.29264</td>
</tr>
<tr>
<td>15</td>
<td>rs11630802</td>
<td>80 028 079</td>
<td>15q25.2</td>
<td>4.88 (\times 10^{-4})</td>
<td>0.001851</td>
<td>0.000706</td>
<td>0.00000</td>
<td>0.00000</td>
<td>NA</td>
</tr>
<tr>
<td>16</td>
<td>rs4843359</td>
<td>84 834 942</td>
<td>16q24.1</td>
<td>0.00277</td>
<td>5.81 (\times 10^{-6})</td>
<td>0.77970</td>
<td>0.88660</td>
<td>0.39124</td>
<td>0.03946</td>
</tr>
<tr>
<td>18</td>
<td>rs9945493</td>
<td>74 604 241</td>
<td>18q23</td>
<td>1.71 (\times 10^{-6})</td>
<td>0.00494</td>
<td>8.97 (\times 10^{-6})</td>
<td>0.91430</td>
<td>0.94111</td>
<td>NA</td>
</tr>
<tr>
<td>20</td>
<td>rs6135782</td>
<td>1 652 114</td>
<td>20q13</td>
<td>6.18 (\times 10^{-4})</td>
<td>0.01629</td>
<td>0.02194</td>
<td>0.70880</td>
<td>0.15025</td>
<td>0.61181</td>
</tr>
<tr>
<td>21</td>
<td>rs2403771</td>
<td>14 650 668</td>
<td>21q11.2</td>
<td>6.20 (\times 10^{-4})</td>
<td>0.00120</td>
<td>0.03668</td>
<td>0.12620</td>
<td>0.47516</td>
<td>0.51669</td>
</tr>
</tbody>
</table>

**Abbreviations:** bp, base pair; Chr, chromosome; Cyto, cytogenetic location; GWAS, genome-wide association study; NA, not available; NIA-LOAD, National Institute on Aging Late-Onset Alzheimer’s Disease; SNP, single-nucleotide polymorphism.

*The SNPs with the most significant \(P\) values \((P < 9 \times 10^{-6})\) in at least 1 of the 3 analyses (overall, \(APOE\) e4 carriers, and \(APOE\) e4 noncarriers) are presented.

Values for the SNPs of interest in the Hispanic GWAS were compared with those in the NIA-LOAD GWAS.

*Family-based allelic association stratified by \(APOE\) status.

*Family-based allelic association taking into account population substructure.

*Flanking SNP within 5 kilobases on either side with a nominal \(P < .05\) in the NIA-LOAD data set.

*These SNPs have 1 or more SNPs with a nominal \(P < .05\) in the NIA-LOAD data set.

**Statistical Analysis**

We conducted single-point allelic association analysis using the Mantel-Haenszel \(x^2\) test statistic, which tests for SNP-disease association conditional on population subcluster estimated from the PLINK analysis described earlier (Table 2). In addition, we performed a multivariate logistic regression analysis, adjusted for age, sex, education, and population stratification, using PLINK (Table 3). For the analysis of all subjects only, we adjusted for the presence or absence of \(APOE\) along with the earlier-mentioned 4 covariates. To determine whether the associations were caused by statistical artifact, we computed the \(P\) value for 1 million replications to derive empirical \(P\) values for the top 23 SNPs that showed the strongest support for association with LOAD. For this purpose, we randomly shuffled affection status for each subject to create the null distribution and assess the likelihood of false-positive results for each SNP.

**Replication Datasets**

We prioritized candidate SNPs by selecting SNPs that had a nominal \(P\) value of \(9 \times 10^{-6}\) or lower. While this cut point does not reach the Bonferroni-corrected genome-wide \(P\) value of .05, this cut point helped us to prioritize SNPs of importance. To determine whether the findings from the Caribbean Hispanic in-
individuals could be replicated in an independent data set, we examined the publicly available GWAS data from the NIA-LOAD study (Wijisman et al, unpublished data [full citation on page 321]) (Table 2). The details of the demographic and clinical characteristics of the NIA-LOAD participants who were included in the GWAS are provided in their report (Wijisman et al, unpublished data [full citation on page 321]). Briefly, the study first examined self-reported European American individuals: 2124 in the published data [full citation on page 321]). Those were augmented with 1186 unrelated individuals from the NIA-LOAD study and 325 individuals from the Rhode Island Institute on Aging Late-Onset Alzheimer’s Disease; OR, odds ratio; SNP, single-nucleotide polymorphism. Abbreviations: Chr, chromosome; CI, confidence interval; Emp, empirical; GWAS, genome-wide association study; MAF, minor allele frequency; NIA-LOAD, National Institute on Aging Late-Onset Alzheimer’s Disease; OR, odds ratio; SNP, single-nucleotide polymorphism.

For the same 23 candidate SNPs, we provide age-, sex-, education-, and population stratification–adjusted ORs from a multivariate logistic regression.

We performed separate analyses focusing on SNPs in the candidate genes that were identified from previous GWAS, including CR1, CLU, PICALM, and BIN1, for the significant genetic associations reported and replicated in 3 previous studies. For these genes, we performed 4 analyses: Mantel-Haenszel χ² test taking into account population stratification, APOE ε4–restricted analysis (ie, restricted to individuals with at least 1 copy of ε4 compared with those without), and Mantel-Haenszel χ² test taking into account the presence or absence of APOE ε4 (Table 4). In addition to those 4 genes, we followed up the novel genetic association identified from the NIA-LOAD GWAS (Wijisman et al, unpublished data [full citation on page 321]). The NIA-LOAD GWAS identified the CUGBP2 gene to be significantly associated with LOAD among a subset of samples with homozygous APOE ε4 carriers. Herein, we evaluated the association using 2 different models to account for its association with the APOE ε4 genotype (Table 4). Under model 1, homozygous APOE ε4 carriers were considered to have the putative genotype and all others did not. Under model 2, homozygous APOE ε4 carriers were considered to have the putative genotype, while homozygous APOE ε3 carriers, the most common isoform, were considered to have a wild type. The remaining subjects were excluded in the analysis.
and 18.2% of the subjects were carriers of an APOE ε4 allele. The mean (SD) age at last examination of the controls was 78.87 (6.4) years. The analysis testing for population stratification revealed that the 1093 Hispanic individuals comprised 658 individuals (60.2%) who were likely to be of European white ancestry, 401 (36.7%) who were likely to be of African ancestry, and 34 (3.1%) who were unrelated to the prior 2 groups and from other Latin American countries (Figure 1).

### STATISTICAL ANALYSIS

None of the SNPs reached genome-wide statistical significance at a nominal P value of $7.97 \times 10^{-8}$ or lower.

The results from the population stratification–adjusted single-point analysis are shown in a Manhattan plot (Figure 2). Twenty-three SNPs had P values less than $9 \times 10^{-6}$ in at least 1 of the 3 analyses, including all combined subjects, carriers of the APOE ε4 allele, and non-carriers of the APOE ε4 allele (Table 2). Of those, the strongest evidence for association was observed for rs9945493 ($P=1.7 \times 10^{-7}$; OR, 0.33; 95% confidence interval, 0.21-0.51) on 18q23. For each SNP, we calculated ORs and 95% confidence intervals as well as empirical P values based on 1 million replicates (Table 3). As observed in other GWAS, ORs ranged from 0.33 for rs9945493 to 1.87 for rs11177750 for all subjects.

We then examined the same 23 SNPs from Table 2 in an independent data set by comparing the results from each of our 3 analyses against data from the NIA-LOAD GWAS, which was restricted to self-reported European American individuals (Wijsman et al, unpublished data [full citation on page 321]). Five SNPs (rs4669573 and rs10197851 on 2p25.1, rs11711889 on 3q25.2, rs1117750 on 7p21.1, and rs7908562 on 10q23.1) from the list of 23 SNPs had P values less than .05 in at least 1 of the 3 analyses in the NIA-LOAD GWAS (Table 2, footnote e); rs4669573 is located within the HPCAL1 (hippocalcin-like 1) gene, and the ODC1 gene is located 100 kilobases away, and rs1117750 and several flanking SNPs that supported allelic association were located within the DGKB (diacylglycerol kinase, β 90 kDa) gene. Lastly, rs7908562 is located proximal to multiple genes, including GHITM (growth hormone inducible transmembrane protein), C10orf99 (chromosome 10 open reading frame 99), PCDH21 (protocadherin 21), LRIT2 (leucine-rich repeat, immunoglobulin-like, and transmembrane domains 2), LRIT1 (leucine-rich repeat, immunoglobulin-like, and transmembrane domains 1), and RGR (retinal G protein-coupled receptor) (eFigure 2).

### Table 4. Replication of Candidate SNPs (CLU, PICALM, and BIN1) From Previous GWAS: Examination in Caribbean Hispanic Individuals*

<table>
<thead>
<tr>
<th>SNP</th>
<th>bp</th>
<th>Population-Stratified Analysis</th>
<th>APOE Stratified</th>
<th>ε4</th>
<th>ε4*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CLU (8p21)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs881146</td>
<td>27500194</td>
<td>0.03740</td>
<td>0.06170</td>
<td>0.00213</td>
<td></td>
</tr>
<tr>
<td>rs70120100</td>
<td>27504646</td>
<td>0.15900</td>
<td>0.12540</td>
<td>0.12210</td>
<td></td>
</tr>
<tr>
<td>rs17057441</td>
<td>27508399</td>
<td>0.04570</td>
<td>0.07790</td>
<td>0.07650</td>
<td></td>
</tr>
<tr>
<td>rs11136000c,d</td>
<td>27520436</td>
<td>0.58190</td>
<td>0.37320</td>
<td>0.62320</td>
<td></td>
</tr>
<tr>
<td><strong>PICALM (11q14)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs17159904</td>
<td>85463935</td>
<td>0.04243</td>
<td>0.04951</td>
<td>0.04270</td>
<td></td>
</tr>
<tr>
<td>rs541458</td>
<td>85465999</td>
<td>0.36300</td>
<td>0.44710</td>
<td>0.29400</td>
<td></td>
</tr>
<tr>
<td>rs543239</td>
<td>85497725</td>
<td>0.72240</td>
<td>0.93160</td>
<td>0.27150</td>
<td></td>
</tr>
<tr>
<td>rs7941541</td>
<td>85536186</td>
<td>0.73180</td>
<td>0.90210</td>
<td>0.27920</td>
<td></td>
</tr>
<tr>
<td>rs3851179c,d</td>
<td>85546228</td>
<td>0.32050</td>
<td>0.46160</td>
<td>0.21020</td>
<td></td>
</tr>
<tr>
<td><strong>BIN1 (2q14)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs10194375</td>
<td>127556011</td>
<td>0.35590</td>
<td>0.38940</td>
<td>0.02755</td>
<td></td>
</tr>
<tr>
<td>rs12486726c,d</td>
<td>127557567</td>
<td>0.04159</td>
<td>0.07147</td>
<td>0.04270</td>
<td></td>
</tr>
<tr>
<td>rs4663098c</td>
<td>127589265</td>
<td>0.71350</td>
<td>0.64590</td>
<td>0.08500</td>
<td></td>
</tr>
<tr>
<td>rs11685593</td>
<td>127604351</td>
<td>0.52240</td>
<td>0.45390</td>
<td>0.02950</td>
<td></td>
</tr>
<tr>
<td>rs7561528</td>
<td>127606107</td>
<td>0.49870</td>
<td>0.46270</td>
<td>0.00500</td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: bp, base pair; GWAS, genome-wide association studies; SNP, single-nucleotide polymorphism.

a Association with SNPs in Caribbean Hispanic samples for the 4 known genes, including CR1, CLU, PICALM, and BIN1. However, none of the SNPs were associated nominally for CR1.

b Allelic association analysis, stratified by ethnicity.

c P values < .05.

d Originally noted as genome-wide significant in Harold et al7 and Lambert et al.9
Association in adjusted analyses. For BIN1 APOE-LOAD in population stratification–adjusted and PICALM individuals, was not associated with LOAD herein. For associated with LOAD in European and American white LOAD in population-stratified analysis and among Table 4, footnote c) was significantly associated with ε APOE We evaluated an interaction model between LOAD, conditional on APOE ε4 carriers. The remaining subjects were excluded from the analysis. bp Indicates base pair. The base pair location on the x-axis is not in scale. In that study, rs201119 in the CUGBP2 –log10 (P) was significantly associated with LOAD only among individuals with a homzygous ε4 genotype (Pnominal=1.52×10−6), but this SNP was not significantly associated with LOAD when all subjects were considered (Pnominal=.726 for allelic association and P=.2607 for genotype association). Because we had a smaller sample size than the NIA-LOAD GWAS, we applied 2 somewhat different models to test whether the allelic association between CUGBP2 and LOAD was restricted to carriers of APOE ε4 and absent in non–APOE ε4 carriers. For this purpose, we performed an interaction model using PLINK in both the Caribbean Hispanic and NIA-LOAD samples. As shown in Figure 3, in the Caribbean Hispanic individuals, we observed a modest interaction between the genotype at rs201119 in the CUGBP2 gene and APOE ε4 genotype (Pnominal=.04898 under model 2). This is the SNP that showed the original allelic association in the NIA-LOAD GWAS samples. For the same SNP, the NIA-LOAD samples had a P value of .00012 under model 1 and .00016 under model 2, supporting the association under our models for both data sets. When we examined all SNPs in CUGBP2 in both data sets, however, we

Figure 2. Manhattan plot of allelic association analysis in a Caribbean Hispanic population. The results of genome-wide association analysis are presented. One single-nucleotide polymorphism has a P value less than 9×10−8 and multiple single-nucleotide polymorphisms have P values less than 9×10−8.

Figure 3. Association between CUGBP2 and late-onset Alzheimer disease (LOAD) among homozygous APOE ε4 carriers in Caribbean Hispanic subjects vs National Institute on Aging Late-Onset Alzheimer’s Disease study European American subjects. Two models were used to examine the relation between CUGBP2 and LOAD, conditional on APOE ε4 status. Model 1 is homozygous APOE ε4 carriers vs others; model 2 is homozygous APOE ε4 carriers vs homozygous APOE ε3 carriers. The remaining subjects were excluded from the analysis. bp Indicates base pair. The base pair location on the x-axis is not in scale.

REPLICATION OF THE PUBLISHED CANDIDATE GENES

For CLU, we observed that rs881146 (Pnominal=.00213; Table 4, footnote c) was significantly associated with LOAD in population-stratified analysis and among APOE ε4 carriers (Table 4). However, rs11136000 in CLU, reported both by Harold et al7 and Lambert et al9 to be associated with LOAD in European and American white individuals, was not associated with LOAD herein. For PICALM, rs17159904 was marginally associated with LOAD in population stratification–adjusted and APOE-adjusted analyses. For BIN1, we observed a positive association in ε4 carriers for rs7561528 (Pnominal=.00536).

GENE × GENE INTERACTION

We evaluated an interaction model between APOE and CUGBP2 to follow up the putative gene × gene interaction finding in the NIA-LOAD study (Wijsman et al, unpublished data [full citation on page 321]) (Figure 3). In that study, rs201119 in the CUGBP2 gene was significantly associated with LOAD only among individuals with a homozygous ε4 genotype (Pnominal=1.52×10−6), but this SNP was not significantly associated with LOAD when all subjects were considered (Pnominal=.726 for allelic association and P=.2607 for genotype association). Because we had a smaller sample size than the NIA-LOAD GWAS, we applied 2 somewhat different models to test whether the allelic association between CUGBP2 and LOAD was restricted to carriers of APOE ε4 and absent in non–APOE ε4 carriers. For this purpose, we performed an interaction model using PLINK in both the Caribbean Hispanic and NIA-LOAD samples. As shown in Figure 3, in the Caribbean Hispanic individuals, we observed a modest interaction between the genotype at rs201119 in the CUGBP2 gene and APOE ε4 genotype (Pnominal=.04898 under model 2). This is the SNP that showed the original allelic association in the NIA-LOAD GWAS samples. For the same SNP, the NIA-LOAD samples had a P value of .00012 under model 1 and .00016 under model 2, supporting the association under our models for both data sets. When we examined all SNPs in CUGBP2 in both data sets, however, we
We report several novel candidate loci that may harbor putative disease variants in Caribbean Hispanic individuals with LOAD and confirmed associations between LOAD and the 4 genes that have been previously reported. These 4 novel loci (5 SNPs) include multiple genes, and further examination is necessary to verify their involvement in LOAD. We replicated the allelic association between LOAD and CUGBP2 in homozygous carriers of the APOE ε4 allele reported by Wijsman and colleagues (unpublished data). This gene was studied because the strongest signal was observed in homozygous ε4 carriers and this region on chromosome 10p14 contains the gene CUGBP2. CUGBP2 has 1 isoform that is expressed predominantly in neurons, with experimental evidence suggesting involvement in apoptosis in the hippocampus. Further, it is involved in posttranscriptional RNA binding activities as well as pre-messenger RNA alternative splicing. Based on structural similarity, it is speculated that this gene may be involved in increasing COX2 messenger RNA. Although the current study does support association with LOAD, the pattern of the associated SNPs differed between the 2 cohorts. The difference in genetic architecture between non-Hispanic and Hispanic populations is the most likely explanation for the fact that the associated SNPs differed between the 2 populations.

We found that the 4 candidate loci that were strongly associated with LOAD and were replicated in the NIA-LOAD cohort are located near genes that could be biologically relevant to LOAD. HPCAL1 on 2p25.1 is a calcium-binding protein expressed in the brain and has been associated with hypertension in Japanese individuals, which in turn is associated with LOAD risk. The region 10q23.1 includes 3 genes that are expressed in the brain and have been reported by Grupe et al, including PCDH21 (believed to be involved in the neuronal maintenance), LRT1, and RGR.

We replicated associations between LOAD and SNPs in 3 of the 4 genes that were previously reported to be significant at the genome-wide level, namely CLU, PICALM, and BIN1. However, the associated SNPs between these candidate genes and LOAD were not necessarily identical in the Caribbean Hispanic individuals compared with a European American data set. Nonetheless, the overall support for the 3 genes is enhanced by the observation that the allelic association extends to an ethnically distinct population.

CLU, believed to be involved in modulation of inflammation and lipid metabolism, was associated with LOAD in carriers of ε4 (P = .00213). More than a decade ago, we examined CLU (also known as APOJ) as a risk factor for LOAD because it shares similar functional roles as APOE, including cholesterol binding and involvement in inflammation or injury. Based on a small set of coding polymorphisms in APOJ, Tycko and colleagues did observe a positive association in 1 homozygous polymorphism, but this association was no longer significant when all subjects with at least 1 copy of the APOE ε4 allele were included. Further, they observed a significant difference in allele frequencies by race, and the present study also shows different linkage disequilibrium patterns between the Caribbean Hispanic individuals and the NIA-LOAD cohorts (eFigure 3). Thus, the inconsistent findings across studies could be attributed to an interaction between APOE and APOJ, small sample size, different distribution of ethnic background in the participants, or any combination of these factors. The present study observed an association between CLU and LOAD in the presence of APOE ε4 (Table 4). This is consistent with the much larger study by Lambert and colleagues but not with the study by Harold et al.

BIN1, a gene expressed in the central nervous system and reported to activate a caspase-independent apoptotic process, was also associated with LOAD in only carriers of ε4 (P = .00536). PICALM is reported to be involved in the neurotransmitter release processes, thereby affecting memory functions.

Together these 3 genes suggest that they contribute to the overall LOAD phenotype. However, the measures of association are unlikely to be consistent across data sets, since in addition to allelic differences among race groups, significant differences in the distribution of vascular and inflammation risk factors can also alter the observed genotype-phenotype relations, even after adjusting for other known risk factors including age, sex, and education.

The current study has some limitations. First, this study, based on a modest sample size of Caribbean Hispanic individuals, does not have power to detect rare variants with weak effects; thus, some risk variants may have been missed. Based on the original GWAS set, the current study has 80% power, genome-wide, to detect alleles with a frequency of 0.35 or higher when the OR is 1.5. When the OR for SNPs is 1.7, this study has 80% power to detect SNPs with an allele frequency of 0.25 or higher. When we combined both Caribbean Hispanic data sets (specifically, one from our GWAS along with the Caribbean Hispanic subset that is part of the NIA-LOAD GWAS), the current study has 80% power genome-wide to detect SNPs with somewhat lower allele frequencies. For a SNP with an OR of 1.5, 80% power can be achieved for SNPs with an allele frequency of 0.3 or higher. For a SNP with an OR of 1.7, 80% power can be achieved for SNPs with an allele frequency of 0.2 or higher. Power calculation was carried out assuming an additive model with SNP minor allele frequency being comparable with the allele frequency of the putative variant (http://pngu.mgh.harvard.edu/~purcell/gpc/cc2.html). Second, independent replication of the candidate SNPs in Caribbean Hispanic individuals who share comparable genetic architecture would have further strengthened the validity of the findings because the likelihood of replicating the same allele within the same SNP would be higher in other ethnic groups. For this reason, we added a small set of Caribbean Hispanic individuals from the NIA-LOAD GWAS data set who were evaluated using the same
diagnostic tools. However, the sample size remained relatively modest. When we evaluated the candidate SNPs in an independent sample of European American individuals with different genetic background (NIA-LOAD GWAS), often allelic associations for the same SNPs were modest, but different SNPs within the gene supported allelic association. However, genetic associations using a cohort with a different ethnic background strengthen the observed association since (1) it is not unexpected to have multiple variants within a gene associated with a disease (eg, PSEN1) and (2) the findings may be generalizable to a wider set of populations. These findings need to be further evaluated using functional genetics approaches to evaluate the validity of observed association.

We used a dense set of SNPs to survey the genome to identify novel loci and to assess support for allelic association with BIN1, CLU, and PICALM. The current cohort extends previous GWAS of non-Hispanic white populations by exploring allelic association in an admixed cohort with a different set of genetic and environmental risk factors. The confirmation in the present study further strengthens the associations between variants in these genes and LOAD. It also supports the role of other genetic (eg, APOE) and environment factors modulating the genetic variant, especially when each variant may only have a small effect size. We also identified novel candidate genes (eg, HPCAL1, DGBK) in a Caribbean Hispanic cohort and replicated the association in an independent ethnically different data set. These genes need to be examined further in independent data sets.

Accepted for Publication: September 2, 2010.

Published Online: November 8, 2010. doi:10.1001/archneur.2010.292

Author Affiliations: Taub Institute for Research on Alzheimer’s Disease and the Aging Brain (Drs Lee, Cheng, Barral, Reitz, Lantigua, and Mayeux), Gertrude H. Sergievsky Center (Drs Lee and Mayeux), and Department of Neurology (Drs Barral, Reitz, and Mayeux), Psychiatry (Dr Mayeux), and Medicine (Dr Lantigua), College of Physicians and Surgeons, and Department of Epidemiology, School of Public Health (Drs Lee and Mayeux), Columbia University, New York, New York; Universidad Tecnológica de Santiago, Santiago, Dominican Republic (Dr Medrano); Department of Internal Medicine, University of Puerto Rico School of Medicine, San Juan, Puerto Rico (Dr Jiménez-Velazquez); Centre for Research in Neurodegenerative Diseases, Departments of Medicine, Laboratory Medicine and Pathobiology, and Medical Biophysics, University of Toronto, and Toronto Western Hospital Research Institute, Toronto, Ontario, Canada (Drs Rogaeva and St. George-Hyslop); and Cambridge Institute for Medical Research and Department of Clinical Neurosciences, University of Cambridge, Cambridge, England (Dr St. George-Hyslop).

Correspondence: Richard Mayeux, MD, MSc, Sergievsky Center, 630 W 168th St, New York, NY 10032 (rpm2@columbia.edu).

Author Contributions: Drs Lee, Cheng, Barral, and Reitz contributed equally to this work. Study concept and design: Lee, Barral, St. George-Hyslop, and Mayeux. Acquisition of data: Medrano, Lantigua, Jiménez-Velazquez, Rogaeva, and Mayeux. Analysis and interpretation of data: Lee, Cheng, Barral, Reitz, and Mayeux. Drafting of the manuscript: Lee, Reitz, Jiménez-Velazquez, and Mayeux. Critical revision of the manuscript for important intellectual content: Lee, Cheng, Barral, Medrano, Lantigua, Rogaeva, St. George-Hyslop, and Mayeux. Statistical analysis: Lee, Cheng, and Barral. Obtained funding: Jiménez-Velazquez, Rogaeva, St. George-Hyslop, and Mayeux. Administrative, technical, and material support: Medrano and Lantigua. Study supervision: Lee, Medrano, Lantigua, and Mayeux.

Financial Disclosure: None reported.

Funding/Support: This work was supported by grants R37-AG15473 and P01-AG07232 from the National Institutes of Health and the National Institute on Aging, the Blanchett Hooker Rockefeller Foundation, Charles S. Robertson Gift from the Banbury Fund (Dr Mayeux), the Canadian Institutes of Health Research, the Howard Hughes Medical Institute, the Canadian Institutes of Health Research–Japan Science and Technology Trust, the Alzheimer Society of Ontario, Wellcome Trust, and Ontario Research Fund (Dr St. George-Hyslop). Dr Reitz was also supported by Paul B. Beeson Career Development Award K23AG034550.


REFERENCES


