Genetic Modifiers of Age at Onset in Carriers of the G206A Mutation in PSEN1 With Familial Alzheimer Disease Among Caribbean Hispanics

Joseph H. Lee, DrPH; Rong Cheng, PhD; Badri Vardarajan, PhD; Rafael Lantigua, MD; Dolly Reyes-Durneyer, BS; Ward Ortmann, PhD; Robert R. Graham, PhD; Tushar Bhangale, PhD; Timothy W. Behrens, MD; Martin Medrano, MD; Ivonne Z. Jiménez-Velázquez, MD; Richard Mayeux, MD, MSc

IMPORTANCE The present study identified potential genetic modifiers that may delay or accelerate age at onset of familial Alzheimer disease (AD) by examining age at onset in PSEN1 mutation carrier families, and further investigation of these modifiers may provide insight into the pathobiology of AD and potential therapeutic measures.

OBJECTIVE To identify genetic variants that modify age at onset of AD.

DESIGN, SETTING, AND PARTICIPANTS Using a subset of Caribbean Hispanic families that carry the PSEN1 p.G206A mutation, we performed a 2-stage genome study. The mutation carrier families from an ongoing genetic study served as a discovery set, and the cohort of those with LOAD served as a confirmation set. To identify candidate loci, we performed linkage analysis using 5 p.G206A carrier families (n = 56), and we also performed whole-exome association analysis using 31 p.G206A carriers from 26 families. To confirm the genetic modifiers identified from the p.G206A carrier families, we analyzed the GWAS data for 2888 elderly individuals with LOAD. All study participants were Caribbean Hispanics.

MAIN OUTCOMES AND MEASURES Age at onset of AD.

RESULTS Linkage analysis of AD identified the strongest linkage support at 4q35 (LOD [logarithm of odds] score, 3.69), and the GWAS of age at onset identified variants on 1p13.1, 2q13, 4q25, and 17p11. In the confirmation stage, gene-wise analysis identified SNX25, PDLIM3, and 3 SH3 domain genes (SORBS2, SH3RF3, and NPHP1) to be significantly associated with LOAD. Subsequent allelic association analysis confirmed SNX25, PDLIM3, and SORB2 as genetic modifiers of age at onset of EOAD and LOAD and provided modest support for SH3RF3 and NPHP1.

CONCLUSIONS AND RELEVANCE Our 2-stage analysis revealed that SNX25, PDLIM3, and SORB2 may serve as genetic modifiers of age at onset in both EOAD and LOAD.

Published online July 27, 2015.

Copyright 2015 American Medical Association. All rights reserved.

Author Affiliations: Author affiliations are listed at the end of this article.

Corresponding Author: Joseph H. Lee, DrPH, Gertrude H. Sergievsky Center, Columbia University, 630 W 168th St, P&S Unit 16, New York, NY 10032 (JHL2@columbia.edu).
In 2001, we identified a p.Gly206Ala (g.44636G>C) variant in PSEN1 (OMIM 104311) in members of 8 Caribbean Hispanic families with early-onset Alzheimer disease (EOAD) from Puerto Rico. Subsequently, Arnold and colleagues also reported this mutation in additional individuals from Puerto Rico but not in other ethnic groups. The mean age at onset among mutation carriers was 55.6 years but was highly variable within families, ranging from 22 to 77 years. The presence of neither an APOE ε4 allele nor any antecedent environmental, health-related, or social factors could explain the differences in the age at onset among mutation carriers, leading us to suspect that other genes were involved in determining when, but not whether, p.G206A carriers develop AD.

Late-onset diseases, including AD, Huntington disease, and Parkinson disease, have been studied for genetic modifiers of age at onset in the presence of a major genetic risk factor (ie, pathogenic mutation), recognizing that there are likely genetic variants that contribute to phenotypic expression. Few studies have reported gene–gene interaction in AD because different individuals are likely to carry different genetic risk factors. Since there are approximately 185 known PSEN1 variants associated with AD and these variants are rare, most studies will not have sufficient statistical power.

We conducted a 2-stage study on a set of families multiply affected by AD with at least one carrier of the p.G206A variant in the PSEN1 gene to identify genetic modifiers of age at onset among carriers. In the discovery stage, we used these genetically homogeneous families in which nearly all carriers of the PSEN1 mutation had p.G206A. Using a subset of carrier family members, we applied both genome-wide linkage and exome sequencing analyses (eFigure 1 in the Supplement). In the confirmation stage, we examined allelic association between age at onset and variants in the same candidate genes in a cohort of the same ancestry with late-onset AD (LOAD). We report that variants in SNX25, PDLIM3, SORBS2, and SH3RF3 were associated with variation in age at onset in families with a copy of PSEN1 p.G206A mutation and in LOAD. In addition, at least one variant in NPH1 was associated with variation in age at onset of AD.

Methods

Discovery Cohort

We conducted a 2-stage study to identify genetic modifiers of age at onset of AD using multiple sets of data in which some were family based and others comprised unrelated individuals (eFigure 1 in the Supplement). The study protocol was approved by the Columbia University Institutional Review Board, and written informed consent was provided by all the participants. Using 56 family members from 5 families with EOAD with the PSEN1 p.G206A mutation, we performed linkage analysis to identify candidate loci. These families represent a subset of PSEN1 carrier families with a wide range of age at onset among affected individuals. Family members resided in the United States, Puerto Rico, Dominican Republic, and countries in South America. Although these families were selected because multiple members had EOAD (age at onset, ≤65 years), many of these families had some family members who had LOAD (age at onset, >65 years). In addition, we performed an exome sequencing experiment and found 31 carriers from 26 families who had multiple affected members and at least 2 who had the PSEN1 p.G206A mutation. Nine of the 31 individuals were also included in the linkage analysis. Given the small sample size of carrier families examined, the main goal of this discovery stage is to prioritize candidate loci.

Confirmation Cohort

For the LOAD validation experiment, we examined 2888 Caribbean Hispanic elderly people who were noncarriers of the PSEN1 p.G206A mutation and were included in separate genome-wide association studies (GWAS) (eTable 1 in the Supplement).

Genetic Analysis

All persons were genotyped using the 6K Illumina linkage panel single-nucleotide polymorphism (SNP) chips (Illumina Inc). This SNP chip was designed to obtain information for linkage signals. Before analysis, we conducted the standard quality control checks, including checking of the reported family relationships using the software PREST, Hardy-Weinberg equilibrium, and genotyping rate. The details of the quality control procedures are presented elsewhere.

Genome-wide Linkage Analysis

To identify regions that may contain variants that could modify age at onset, linkage analyses were performed using the 6K linkage panel SNP data from 56 individuals in 5 families. We examined AD and age at onset as phenotypes. We reasoned that variants that lead to early age at onset are likely to increase the risk of AD, whereas variants that delay age at onset may reduce the risk. For AD phenotype, single-point analysis assumed an autosomal dominant mode of inheritance and used the affecteds-only approach in PSEUDOMARKER. For age-at-onset phenotype, we used age at onset for affected individuals and age at last examination for unaffected individuals. We then applied the multipoint variance component method to identify candidate loci as implemented in SOLAR statistical software. This variance component model was adjusted for sex, APOE ε4, and AD status.

To prioritize candidate genes under the linkage peak at 4q35.1, the top linkage peak, we screened candidate genes using CANDID, version 1.1. Higher weights were given to logarithm of odds (LOD) score (weight, 2) and conservation (weight, 2) compared with literature (which included the keywords Alzheimer disease, neurodegeneration, early onset, and beta amyloid) (weight, 1) and gene expression (weight, 1).

Exome Sequencing

Whole-exome capture libraries were constructed from DNA from whole blood of the samples after sample shearing, end repair, phosphorylation, and ligation to bar-coded sequencing adaptors (eFigure 2 in the Supplement). The ligated DNA was subjected to exonic hybrid capture using the Nimble exome capture array (http://www.nimblegen.com/sequakep2). Samples were multiplexed and sequenced on multiple Illumina HiSeq
flow Mapping, Variant Calling, and Downstream Bioinformatics Analyses
Prealignment quality control of the read data to identify failed runs and lanes was performed using the R BioConductor package: ShortRead (http://www.bioconductor.org/). Sequence alignment was performed using the aligner BWA. Polymerase chain reaction duplicates (reads) were removed using Picard tools. Base quality recalibration and realignment around indels were performed using the Genome Analysis Toolkit (https://www.broadinstitute.org/gatk/). Multisample variant calling and quality control of the call were then performed using the Genome Analysis Toolkit. On the basis of the hypothesis that the mutation underlying this rarer form of EOAD was not present in the general population, SNPs identified in the 1000 Human Genomes (HG) project (version 2010-11; http://www.1000genomes.org) or in dbSNP (build 134; http://www.ncbi.nlm.nih.gov/projects/SNP) were removed. Exonic coding variants were identified and annotated by ANNOVAR (http://annovar.openbioinformatics.org/en/latest/). Non-exonic and synonymous variants were filtered from the variant list using predictions from SIFT (Sorting Intolerant From Tolerant) software, version 4.0 (http://sift.jcvi.org).

Genome-wide Association Analysis
Using 31 individuals with exome sequencing data, we examined candidate genes under the top linkage peak and the rest of the exome by applying a mixed linear model adjusting for sex, APOE genotype, affection status, and kinship coefficient. The kinship coefficient matrix using the R functions21 was included in the model to take into account relatedness among family members.

Replication Using the LOAD Hispanic GWAS Data Set: Genewise Analysis
To determine whether candidate genes identified from the EOAD cohort were similarly associated with age at onset of AD in the LOAD cohort of the same ethnic background, we performed genewise association analysis using family-based SKAT software (FAMSKAT),22 adjusting for sex, AD status, and APOE ε4. This approach was necessary because (1) multiple variants within one gene had modest effects on age at onset and (2) most of the variants identified from exome sequencing were absent in the GWAS data of Caribbean Hispanics.

Imputation of the LOAD Hispanic GWAS Data Set
Imputations of 5 genes (SNX25, PDLIM3, SORBS2, SH3BP3, NPHS1) were performed based on cosmopolitan phased haplotypes of 1000 HG, version 2010-11 (data freeze, 2012-03-04 haplotypes; http://csg.sph.umich.edu/abecasis/MaCH/download/1000G.2012-03-14.html). PLINK, version 1.07 (http://pngu.mgh.harvard.edu/~purcell/plink/) was used for imputation. Before imputing, a number of filters were implemented in the Hispanic GWAS genotypic data by removing markers that had a minor allele frequency less than 1%, an Hardy-Weinberg equilibrium $P < 10^{-6}$, Hispanic GWAS SNP alleles mismatched with those of 1000 HG project and not present in the 1000 HG panel, and flipping of any SNP when appropriate to the forward strand. A total of 348,924 SNPs for 2 regions of chromosome 2 (175,288 SNPs, approximately 14 Mbp) and chromosome 4 (173,636 SNPs, approximately 10 Mbp) were imputed. For single-variant and single-trait association with imputed genotypes, 2 additional filters were implemented: minor allele frequency greater than 1% and an information score greater than 0.8 (a quality score from the imputation), which reduced the analysis to 2636 variants within the 3 genes.

Association Analysis and Meta-analysis
We applied the same multivariate linear mixed model that adjusted for sex, AD status, and APOE ε423 as above to test the allelic association between age at onset and variants in LOAD. We then conducted a meta-analysis to assess whether the SNPs in the 3 candidate genes were significantly associated with the LOAD GWAS data sets. For this purpose, we estimated the meta-analysis of $P$ values from 4 studies as implemented in the METAL (http://csg.sph.umich.edu/abecasis/metal/), which estimates a single summary $P$ value across 3 data sets. An overall $z$ statistic and $P$ value were calculated while taking into account the number of individuals examined in each study.

Results

Study Participants
Table 1 details the 3 sets of Caribbean Hispanics investigated: (1) families who have multiple affected members and at least one carrying the PSEN1 p.G206A mutation were analyzed using linkage analysis; (2) carriers of the PSEN1 p.G206A mutation from families who have multiple affected members were analyzed using exome sequencing; and (3) 3 confirmatory GWAS data sets of Caribbean Hispanics primarily composed of LOAD were studied,15,23 which were analyzed using a linear mixed model to assess allelic association. All 3 data sets had both family members and unrelated individuals, and appropriate statistical methods were applied to account for nonindependence.

Linkage Analysis Set
Of 56 family members included in the linkage analysis, 21 (37.5%) were affected, and 32 (57.1%) were unaffected. A total of 37 (66.1%) of the family members were women, the mean (SD) age at onset of AD was 57.2 (7.0) years (range, 43.0-73.0 years), and the mean educational level was 9.1 years. Most mutation carriers reported
Puerto Rico as the country of origin, whereas the others were from the Dominican Republic.

**Caribbean Hispanic GWAS Set**

Among 2888 individuals in the confirmation set, 1473 (51.0%) of the individuals were affected with AD, and 1918 (66.4%) were women. Most affected individuals had LOAD, and their mean age at onset was 74.0 years (range, 30.0-105.0 years). The proportion of APOE ε4 carriers (23.0%) was comparable to those with EOAD family members.

**Genome-wide Linkage Analysis of Families Carrying the PSEN1 p.G206A Mutation**

For the purpose of gene discovery, we conducted genome-wide linkage analyses of AD and age at onset using 5 multiplex families with variable age at onset. The linkage analysis of AD iden-

### Table 1. Demographic and Clinical Characteristics of Sequenced Participants and Family Members*

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Genotyped (n = 56 [5 Families])</th>
<th>Sequenced (n = 31 [26 Families])</th>
<th>Combined GWAS (N = 2888)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affection status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affected</td>
<td>21 (37.5)</td>
<td>26 (83.9)</td>
<td>1473 (51.0)</td>
</tr>
<tr>
<td>Unaffected</td>
<td>32 (57.1)</td>
<td>5 (16.1)</td>
<td>1408 (48.8)</td>
</tr>
<tr>
<td>Unknown</td>
<td>3 (5.4)</td>
<td>0</td>
<td>6 (0.2)</td>
</tr>
<tr>
<td>Sex, male:female</td>
<td>19:37 (33.9:66.1)</td>
<td>12:19 (38.7:61.3)</td>
<td>970:1918 (33.6:66.4)</td>
</tr>
<tr>
<td>Age at onset or age at last examination, mean (SD), y</td>
<td>57.2 (7.0)</td>
<td>57.9 (9.9)</td>
<td>74.0 (10.0)</td>
</tr>
<tr>
<td>Educational level, mean (SD), y</td>
<td>10.0 (4.6)</td>
<td>9.1 (5.2)</td>
<td>5.7 (4.7)</td>
</tr>
<tr>
<td>APOE frequency</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E4</td>
<td>23 (20.5)</td>
<td>19 (30.6)</td>
<td>1238 (23.0)</td>
</tr>
<tr>
<td>E3</td>
<td>85 (75.9)</td>
<td>43 (69.4)</td>
<td>3849 (71.5)</td>
</tr>
<tr>
<td>E2</td>
<td>4 (3.6)</td>
<td>0 (0.0)</td>
<td>293 (5.5)</td>
</tr>
</tbody>
</table>

Abbreviation: GWAS, genome-wide association study.

* Data are presented as number (percentage) of participants unless otherwise indicated. The combined GWAS data set has a total of 2324 unrelated individuals (2321 families and 3 individuals who married into the families).

**Figure. Results From Linkage and Association Analysis for 4q35**

The bars represent logarithm of odds (LOD) scores from single-point linkage analysis, and points without the bar represent -log10 P values from association analysis. Approximate locations for relevant genes as well as physical locations on chromosome 4 (chr4) are presented at the bottom of the figure.
was observed at rs906815 in NPHP1 at 2q13 (P = 4.51 × 10⁻⁶). Specifically, individuals carrying the rare A allele for rs906815 had an age at onset approximately 11.7 years earlier for having one copy of the A allele compared with those who do not, thereby leading to nearly a difference of 2 decades in age at onset (ie, 44 vs 64.17). In addition, the effect variants in SNX25 and SH3RF3 were associated with a delay in age at onset by 8.8 and 9.3 years, respectively, whereas the effect variant in PDLIM3 was associated with an earlier onset by 12 years. In addition, rs906815 on 2q13 (P = 4.5 × 10⁻⁶) and rs3026115 on 17p13 (P = 7.8 × 10⁻⁵) had SNPs with strong support for allelic association. To be conservative, we performed one additional analysis of age at onset of AD in which we restricted the analysis to affected individuals only. This analysis yielded comparable results as the survival analysis approach. For example, rs906815 in NPHP1 had a P value of 4.13 × 10⁻⁶, and other SNPs that were significant in the earlier analysis remained statistically significant, but the P values were somewhat weakened (eTable 2 in the Supplement).

Replication in Individuals With LOAD: Genewise Analysis
To determine whether the variants discovered from the PSEN1 p.G206A carrier families were likely to be associated with variation in age at onset in LOAD, we first conducted genome-wide association analysis using FAMSKAT because the Hispanic GWAS data sets did not include the variants identified from exome sequencing of the discovery set. We confirmed that SNX25 (P = 1.26 × 10⁻¹⁵), PDLIM3 (P = 1.17 × 10⁻¹³), SORBS2 (P = 3.8 × 10⁻⁶), NPHP1 (P = 5.47 × 10⁻⁵), and SH3RF3 (P = 2.7 × 10⁻¹³) were associated with age at onset in Caribbean Hispanics with LOAD, further supporting that these genes may be involved functionally in altering phenotypic expression in AD.

Replication in Individuals With LOAD: SNP-wise Analysis
To further explore whether allelic association persists within those candidate genes in LOAD, we examined SNPs from the 3 GWAS data sets of Caribbean Hispanic ancestry, focusing on the SNPs in the same linkage disequilibrium block as the original discovery variants. Because the GWAS data sets did not have genotypes for many of the variants discovered from exome se-

### Table 2. Summary of Linkage, Allelic, and Genewise Association Analyses in EOAD and LOAD

<table>
<thead>
<tr>
<th>Gene</th>
<th>SNP Location, bp</th>
<th>β</th>
<th>P Value</th>
<th>No. of SNPs*</th>
<th>P Value for Genewise Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Linkage Signal</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4q35</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>STOX2</td>
<td>rs28411264</td>
<td>184908447</td>
<td>13.01</td>
<td>.03851</td>
<td>12</td>
</tr>
<tr>
<td>ENPP6</td>
<td>rs13147555</td>
<td>185074686</td>
<td>-6.06</td>
<td>.03024</td>
<td>28</td>
</tr>
<tr>
<td>MLF1IP</td>
<td>rs1401359</td>
<td>185640216</td>
<td>-5.11</td>
<td>.09064</td>
<td>58</td>
</tr>
<tr>
<td>SNX25</td>
<td>rs11730401</td>
<td>186268129</td>
<td>8.75</td>
<td>.0004</td>
<td>17</td>
</tr>
<tr>
<td>PDLIM3</td>
<td>rs28522047</td>
<td>186446018</td>
<td>-11.97</td>
<td>.1218</td>
<td>13</td>
</tr>
<tr>
<td>SORBS2</td>
<td>rs13130022</td>
<td>186678161</td>
<td>11.11</td>
<td>.0013</td>
<td>143</td>
</tr>
<tr>
<td>FAM149A</td>
<td>rs2276920</td>
<td>187078938</td>
<td>7.11</td>
<td>.01671</td>
<td>14</td>
</tr>
<tr>
<td><strong>Association Signal</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2q13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SH3RF3</td>
<td>rs6542814</td>
<td>109928460</td>
<td>9.30</td>
<td>.001</td>
<td>157</td>
</tr>
<tr>
<td>NPHP1</td>
<td>rs906815</td>
<td>110942496</td>
<td>-11.69</td>
<td>4.51 × 10⁻⁶</td>
<td>12</td>
</tr>
</tbody>
</table>

Abbreviations: bp, base pair; EOAD, early-onset Alzheimer disease; GWAS, genome-wide association study; LOAD, late-onset Alzheimer disease; SNP, single-nucleotide polymorphism.

* The SNP with the strongest P value from the exome sequencing experiment is given.
quencing, we performed imputation using PLINK\textsuperscript{24} and then used both genotype and imputed SNPs for allelic association. We limited the use of imputed SNPs to those with an information score greater than 0.8 by PLINK analysis. As shown in eFigure 3 in the Supplement, pairwise D′ values across the chromosomal regions encompassing the discovery variants in candidate loci were high and were comparable across the 3 GWAS data sets.

Table 3 indicates that multiple genotyped or imputed data in the candidate genes were associated with age at onset in LOAD. In this confirmation analysis, 14 SNPs in SNX25 were associated with age at onset at \( P < .05 \), and variants were associated with 1 to 6 years of difference in age at onset in LOAD. For \textit{PDLIM3}, 3 SNPs were associated with variation in age at onset and were associated with a difference of 2.6 to 3.6 years in mean age at onset. For \textit{SORBS2}, 11 of 1246 SNPs or indels flanking the original discovery variant were significantly associated with age at onset of LOAD at \( P < .05 \), and the variants were associated with 1 to 4.6 years of mean age difference in onset. It is of interest to note that, for \textit{SORBS2}, the same allele was associated with age at onset in all 3 GWAS data sets, and the allele frequency for the associated SNP was consistent across data sets, ranging from 0.032 to 0.222 with a median of 0.194. For \textit{SH3RF3}, 5 SNPs were associated with age at onset at \( P < .05 \) in LOAD, and the difference in mean age at onset ranged from 1 to 2.9 years. For \textit{NPHPI}, only one SNP was associated with age at onset.

Table 4 summarizes reanalysis of the 10 SNPs from the 2 candidate genes (\textit{SORBS2} and \textit{SH3RF3}) using genotyped data. The rationale for genotyping these subsets of SNPs from Table 3 was that the information score for the imputed SNPs in those 2 genes did not reach the satisfactory quality control threshold of 0.8 according to our PLINK analysis; thus, the results from the meta-analysis based on smaller data sets are likely to be less reliable when compared with other SNPs with high information scores. When genotyped data were used, rs72939527 in \textit{SH3RF3} had the strongest support for allelic association (\( P = .007 \)). The T allele in rs72939527 was associated with a 2.7-year earlier age at onset compared with the reference allele, whereas other variants did not reach statistical significance (\( P = .22-.47 \)).

Discussion

We focused on founder mutation \textit{PSEN1} p.G206A with high penetrance to identify genetic modifiers of age at onset of EOAD and LOAD and report that variants in \textit{SNX25}, \textit{PDLIM3}, \textit{SORBS2}, \textit{SH3RF3}, and \textit{NPHPI} may contribute to variation in age at onset in EOAD and LOAD. This study of a founder mutation allows identification of a second locus that may alter the effect of the highly penetrant, nonsynonymous \textit{PSEN1} p.G206A variant. This approach overcomes some of the limitations that are common in studies that examine gene–gene interaction in variants with modest to weak effect sizes. Furthermore, changes in age at onset by genotype among \textit{PSEN1} p.G206A carrier families ranged from 10 to 20 years, whereas those among individuals with LOAD ranged from 1 to 6.2 years. This finding suggests that the difference in age at onset by genotype is likely to be smaller in noncarriers compared with G206A carriers.

Familial EOAD is caused by mutations in 3 genes: \textit{APP}, \textit{PSEN1}, and \textit{PSEN2}.\textsuperscript{25-28} Support clearly exists for genetic modifiers because age at onset varies widely for carriers of \textit{APP}, \textit{PSEN2}, and \textit{PSEN1}. To date, only a few studies\textsuperscript{3,4,11} have reported modifiers. Investigators have attributed observed variable age at onset in \textit{PSEN2} to the APOE ε4 allele.\textsuperscript{3,4} For \textit{PSEN1}, Vélez and colleagues\textsuperscript{11} conducted a pooling- and bootstrap-based GWAS to identify multiple modifiers. Among those, they observed that rs10173717 in \textit{NPHPI} may alter age at onset in \textit{PSEN1} p.Glu280Ala carriers. However, rs10173717 reported by Vélez et al\textsuperscript{11} was located 40.5 kilobases (kb) away from our own discovery signal rs906815 and 76.1 kb away from our GWAS signal rs17842680. The authors did not report the difference in age at onset by genotype.

In this study, 5 promising genetic modifiers were identified among families multiply affected by LOAD: \textit{SNX25}, \textit{PDLIM3}, \textit{SORBS2}, \textit{SH3RF3}, and \textit{NPHPI}. \textit{SNX25} has been implicated in regulating endosome sorting and signaling.\textsuperscript{29,30} Given its potential role in regulating membrane protein trafficking excess levels of amyloid β in individuals with the \textit{PSEN1} p.G206A variant, this gene may be a biologically relevant modifier. This hypothesis is further supported by an earlier report\textsuperscript{31} that found that \textit{SNX3}, a gene in the same family, was associated with \textit{AD}, and \textit{SNX3} regulates and interacts with the retromer membrane. \textit{PDLIM3} is reported to be associated with diabetes and cardiomyopathy.\textsuperscript{32} Although there have been reports of insulin-related genes, little is known about the role of \textit{PDLIM3} in dementia to date. Interestingly, however, the PDZ domain is regulated by \textit{SNX27} retromer protein, and deficiencies in retromers have been associated with \textit{AD}.\textsuperscript{33}

The remaining 3 candidate genes encode proteins with a functional domain, \textit{SH3}. These evolutionarily conserved domains are defined by sequence homology and are involved in protein–protein interaction.\textsuperscript{34,35} \textit{SH3} domains are ubiquitous intracellular protein modules; structural and functional studies\textsuperscript{34-37} reveal how they interact with their praline-rich ligands to promote the formation of specific protein aggregates. The \textit{SH3} domain is related to the WW and PTB domains\textsuperscript{37} and may interact with \textit{APP} to modulate the degradation of \textit{APP}.\textsuperscript{35} The \textit{SORBS2} gene—an \textit{SH3}-binding domain gene located in 4q35.1 under the strongest linkage peak—transcribes a brain-specific splice variant known as nARGBP2.\textsuperscript{38} This variant was reported to influence the integrity of f-actin in the dendritic spines of neurons. Cestra and colleagues\textsuperscript{39} observed that overexpression of nARGBP2 in mice caused aggregation of f-actin bundles in dendritic spines and may influence phenotypic expression. Because alteration of synaptic shape may be associated with the biological progression of \textit{AD}, variants in \textit{SORBS2} may influence variation in age at onset by affecting dendritic spine morphologic and subsequent neurodegenerative processes.

\textit{SH3RF3} is an important paralog of \textit{SORBS2}. Because paralogs arise from gene duplication within the same species and we found allelic association between age at onset and \textit{SORBS2} as well as \textit{SH3RF3}, \textit{SH3RF3} is a biologically plausible gene. In the present data set, the rare homozygous genotype carriers had age at onset approximately 20 years earlier than those with the wild-type homozygous genotype under the model that ad-
adjusted for sex, AD, APOE ε4, and kinship coefficient. These results suggest that some individuals harbor a variant in SH3RF3 that may modify the effect of the PSEN1 p.G206A mutation on the disease phenotype. We note that the SH3 domain is found in proteins of signaling pathways regulating the cytoskeleton. They also regulate the activity state of adaptor proteins.
of Alzheimer Disease by Genotype studies41-43 have found that when 2 cohorts of different ethnic populations. Thus, the identified variantsthemselvesmayhavein Caribbean Hispanics but not in the general US netic variants identified from this study mayberelativelycom- ported age at onset is an approximation. In addition, the ge- development of AD using sensitive biomarkers. Thus, re- vidual is followed up longitudinally from preclinical stage to is difficult to detect prodromal symptoms unless each indivi- dual is followed up longitudinally from preclinical stage to development of AD using sensitive biomarkers. Thus, re- ported age at onset is an approximation. In addition, the ge- netic variants identified from this study may be relatively common in Caribbean Hispanics but not in the general US populations. Thus, the identified variants themselves may have limited value in other ethnic cohorts. However, our earlier studies41-43 have found that when 2 cohorts of different eth-

### Table 4. Reanalysis of SH3RF3 and SORBS2 in Late-Onset Alzheimer Disease Using Genotype Data and Difference in Age at Onset of Alzheimer Disease by Genotype

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>SNP</th>
<th>Location, bp (HG19)</th>
<th>A1</th>
<th>No. of Genotyped Persons</th>
<th>Age-at-Onset Phenotype</th>
<th>Alzheimer Disease Phenotype</th>
<th>Onset Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>rs72935927</td>
<td>109 917 895 T</td>
<td>1705</td>
<td>-2.863</td>
<td>-2.692</td>
<td>.007</td>
<td>1.123</td>
</tr>
<tr>
<td>2</td>
<td>rs6542814</td>
<td>109 928 460 A</td>
<td>1697</td>
<td>0.374</td>
<td>1.231</td>
<td>.22</td>
<td>1.092</td>
</tr>
<tr>
<td>2</td>
<td>rs11679797</td>
<td>109 973 646 G</td>
<td>1703</td>
<td>-1.007</td>
<td>-1.032</td>
<td>.30</td>
<td>1.091</td>
</tr>
<tr>
<td>2</td>
<td>rs13013000</td>
<td>109 982 224 T</td>
<td>1697</td>
<td>-0.695</td>
<td>-0.729</td>
<td>.47</td>
<td>1.114</td>
</tr>
<tr>
<td>2</td>
<td>rs11674377</td>
<td>109 984 093 A</td>
<td>1699</td>
<td>-1.037</td>
<td>-1.073</td>
<td>.28</td>
<td>1.100</td>
</tr>
<tr>
<td>2</td>
<td>rs17842680</td>
<td>110 906 930 G</td>
<td>1700</td>
<td>0.686</td>
<td>1.054</td>
<td>.29</td>
<td>1.028</td>
</tr>
<tr>
<td>2</td>
<td>rs906815</td>
<td>110 942 496 A</td>
<td>1702</td>
<td>-0.329</td>
<td>-0.105</td>
<td>.31</td>
<td>1.116</td>
</tr>
</tbody>
</table>

**SORBS2**

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>SNP</th>
<th>Location, bp (HG19)</th>
<th>A1</th>
<th>No. of Genotyped Persons</th>
<th>Age-at-Onset Phenotype</th>
<th>Alzheimer Disease Phenotype</th>
<th>Onset Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>rs10009306</td>
<td>186 663 793 C</td>
<td>1700</td>
<td>0.103</td>
<td>0.244</td>
<td>.81</td>
<td>1.077</td>
</tr>
<tr>
<td>4</td>
<td>rs1021678</td>
<td>186 678 768 A</td>
<td>1700</td>
<td>0.272</td>
<td>0.700</td>
<td>.48</td>
<td>1.111</td>
</tr>
</tbody>
</table>

Abbreviations: A1, allele 1; bp, base pair; HG19, human genome assembly 19; SNP, single-nucleotide polymorphism.

* rs113309962 (109987615 bp) was homozygous.

Conclusions

This study of carriers of the PSEN1 p.G206A mutation has identified variants in SNX25, PDLIM3, and a family of SH3 domain genes (SORBS2, SH3RF3, NPHP1) that may modify age at onset of EOAD. Furthermore, variants in SORBS2 consistently were associated with delayed onset in EOAD and LOAD, although less profoundly so in LOAD. On the other hand, the discovery variant and flanking variants in SH3RF3 and NPHP1 were significantly associated with EOAD, but their effect size was modest with LOAD. One possible explanation for lack of association in the present data set may be that the effect size for SH3RF3 and NPHP1 may be relatively modest, and a larger study may be needed to detect significant association. The present findings are worthy of further investigation to understand the pathobiology of AD.
Genetic VariantsofAlzheimer Disease Onset


