Genome-Wide Association Study of Incident Dementia in a Community-Based Sample of Older Subjects


*Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA
bDepartment of Neurology, School of Medicine, University of Pittsburgh, Pittsburgh, PA, USA
cDepartment of Neurology, College of Medicine, University of Florida, FL, USA

Received 11 May 2022
Pre-press 9 June 2022

Abstract
Background: Alzheimer’s disease (AD) is a complex disease influenced by the environment and genetics; however, much of the genetic component remains unaccounted for.
Objective: The purpose of this work was to use genome-wide association analyses to detect genetic associations with incident AD in a sample of older adults aged 75 and above.
Methods: We performed a genome-wide association study (GWAS) on genome-wide genotyped and imputed data (14,072,053 variants) on the Gingko Evaluation of Memory (GEM) study sample consisting of 424 incident dementia (mean age = 84.46 ± 3.91) and 2,206 non-demented (mean age = 84.55 ± 3.23) subjects.
Results: The established association of APOE*4 carriers with AD was confirmed in this community-based sample of older subjects (odds ratio (OR) = 2.22; p = 9.36E-14) and was stronger in females (OR = 2.72; p = 1.74E-10) than in males (OR = 1.88; p = 2.43E-05). We observed a novel genome-wide significant (GWS) locus on chromosome 12 near ncRNA LOC105369711/rs148377161 (OR = 3.31; p = 1.66E-08). In addition, sex-stratified analyses identified two novel associations in males: one near ncRNA LOC729987/rs140076909 on chromosome 1 (OR = 4.51; p = 3.72E-08) and the other approaching GWS near ncRNA LOC105375138/rs117803234 on chromosome 7 (OR = 3.76; p = 6.93E-08).
Conclusion: The use of community-based samples of older individuals and incident dementia as a phenotype may be a helpful approach for the identification of novel genes for AD, which may not be detected in standard case-control studies. Replication of these signals and further studies of these regions and genes will help to provide a clearer picture for their role in AD.

Keywords: Alzheimer’s disease, genome-wide association study (GWAS), incident dementia, non-coding RNA genes

INTRODUCTION

Dementia is a broad term describing conditions that affect a person’s ability to remember and make decisions, and can affect behavior and personality. Alzheimer’s disease (AD) is the most common form of dementia, making up 60–80% of dementia cases and predominantly affecting the elderly [1]. In the United States, an estimated 6.2 million people 65 or older were affected in 2021, and due to demographic changes this number is estimated to rise to approximately 14 million by 2060 [1]. AD is a complex disease influenced by genetics...
as well as the environment, with heritability estimated to be between 58–70% [2, 3]. Currently, about 31% of genetic variance is explained by common single-nucleotide polymorphisms (SNPs) from genome-wide association studies (GWAS) [4]. Further studies are needed to account for this unexplained genetic component to gain better understanding of the disease. Most AD genetic studies have been done on highly selected cases derived from AD research centers having moderate to severe AD [5]. On the other hand, community-based cohorts generally have cases with mild dementias, which may help to identify novel genes not detectable in standard case-control studies. Thus, the objective of this work was to use genome-wide association analyses to detect novel genetic associations with incident AD in a community-based sample of older adults (age range 75 to 99 years) derived from the Ginkgo Evaluation of Memory (GEM) study.

**METHODS**

The study subjects were derived from the GEM study, the purpose of which was to test the effect of *Ginkgo biloba* on the development of incident dementia in community populations which were cognitively normal or had mild cognitive impairment (MCI) [6]. Subjects were recruited from nearby communities at four clinical sites: University of Pittsburgh, University of California-Davis, John Hopkins University, and Wake Forest University [6]. The study participants were screened to determine if they met inclusion or exclusion criteria. Cognitive assessment was done initially with a telephone interview that if passed would result in a scheduled screening visit. The Modified Mini-Mental State Examination (3MSE) [7] was used to further assess cognitive function before administering a more detailed neuropsychological battery. The neuropsychological battery consisted of multiple tests designed to assess the following cognitive domains: estimated verbal IQ, memory, construction, language, psychomotor speed, attention, and executive function [8].

Individuals who already showed signs of dementia or those who had other neurological conditions were excluded from the study; those who showed no signs of dementia or showed signs of MCI were included. Subjects who were taking over-the-counter cognitive enhancers, cholinesterase inhibitors, were not willing to restrict their vitamin E intake, or who were taking over-the-counter *G. biloba* were excluded. Those with an increased risk of bleeding either due to a disorder, low platelets, or use of anticoagulant medication were excluded due to potential increased bleeding risk from use of *G. biloba*. Details of other exclusion criteria are found in DeKosky et al. [9].

A total of 3,069 subjects (95.5% Whites and 4.5% non-Whites) were recruited. DNA was available on 2,737 subjects (96.1% White and 3.9% non-White), who were genotyped using the Illumina Infinium Multi-Ethnic Global-8 v1.0 chip containing 1,779,819 single-nucleotide polymorphisms (SNPs). Fifteen samples were excluded for having a call rate below 95%. One sample failed in the genotyping chip analysis. Imputation was done using the Haplotype Reference Consortium (HRC) panel on the Michigan Imputation Server (https://imputationserver.sph.umich.edu/). After imputation, the total number of SNPs increased to 14,072,053. Genotypes for the APOE/rs429358 (APOE*4) and APOE/rs7412 (APOE*2) SNPs were determined using TaqMan genotyping assays. Because of the strong linkage disequilibrium (LD) between the two sites, this is also treated as a three-allele APOE polymorphism: APOE*2, APOE*3, and APOE*4, resulting in six genotypes (2/2, 2/3, 2/4, 3/3, 3/4, 4/4) [5].

Principal components (PCs) of ancestry were calculated using variants with minor allele frequency (MAF) of >0.05. A sliding window approach was used with a 2,000 bp window size shifted every 200 variants to estimate correlation. This was done to avoid using highly correlated variants to estimate ethnicity structure of the cohort. Variants were identified and removed from principal component analysis with maximum likelihood phasing based r² > 0.05. The first four principal components generated from this analysis were used as covariates in the GWAS analyses.

Because of the small number of non-Whites, only White individuals (n = 2,630; age range 75 to 99 years) were used for genetic analyses (Table 1). Logistic regression was used to associate SNPs with dementia risk. Two association analyses were done based on the inclusion of: 1) all incident cases including any form of dementia, and 2) incident cases including only AD Dementia. The Any Dementia analysis used 424 incident cases (386 AD Dementia and 38 non-AD Dementia) and 2,206 controls with a total of 2,630 individuals (Table 1). The AD Dementia analysis used 386 incident AD cases and 2,206 controls for a total of 2,592 individuals (Table 1).

Both analyses used sex, education, and age-at-onset (AAO) for cases, age at last follow-up for controls, and the first four PCs of ancestry as covariates.
Gender-stratified (male and female) and APOE genotype-stratified (E*4 and non-E*4 carriers) analyses were done separately for Any Dementia and AD Dementia. The analysis of males with Any Dementia consisted of 204 with AD dementia, 17 with other dementia and 1,228 controls. The analysis of females with Any Dementia consisted of 182 with AD Dementia, 21 with other dementia, and 978 controls. The analysis of males with AD Dementia included 204 with AD and 1,228 controls. The analysis of females with AD Dementia consisted of 182 with AD and 978 controls. Covariates used for this sex-stratified analysis included the first four PCs of ancestry, AAO for cases, age at last follow-up for controls, and education. APOE genotype-stratified analyses were carried out among APOE*4 and non-APOE*4 carriers. The analysis of APOE*4 carriers with Any Dementia consisted of 150 with AD Dementia and 9 with other dementia and 458 controls. The analysis of non-APOE*4 carriers with Any Dementia consisted of 232 with AD Dementia and 29 with other dementia and 1,739 controls. The analysis of APOE*4 carriers with AD Dementia consisted of 150 with AD Dementia and 458 controls. The analysis of non-APOE*4 carriers with AD Dementia consisted of 232 with AD and 1,739 controls. Covariates used for these analyses were the first four PCs of ancestry, sex, AAO for cases, age at last follow-up for controls, and education.

These association analyses were done using Plink [10]. Manhattan and QQ plots were generated in R using the qqman package [11]. SNPs that did not meet Hardy-Weinberg Equilibrium (HWE test \( p \leq 1E-06 \)) were removed. The genome-wide significance threshold was set at \( p < 5E-08 \), while the suggestive significance threshold was set at \( p < 1E-05 \). Regional association plots were generated using LocusZoom [12] based on GRCh37 assembly.

Functional annotation for newly identified SNPs and for those in LD was done by looking at those SNPs in the Genotype-Tissue Expression project (GTEx) v8 database to check for any single-tissue expression quantitative trait loci (eQTLs). These SNPs were also checked in the RegulomeDB database to assess their potential regulatory function. Combined Annotation Dependent Depletion (CADD) scores designed to predict the deleteriousness of a variant were also obtained for these SNPs.

### RESULTS

**APOE genotype distribution**

Of the 2,617 White samples successfully genotyped for APOE/rs429358 (APOE*4) and APOE/rs7412 (APOE*2) SNPs, 2,579 were used in the AD Dementia case-control analysis and 2,617 individuals in the Any Dementia case-control analysis. The distribution of the six APOE genotypes derived from the two SNPs was in HWE in both the AD Dementia and Any Dementia case-control analyses (Table 2; Supplementary Table 1). As expected, the overall distribution of APOE genotypes was significantly different between AD Dementia and cognitively normal groups (\( p = 1E-05 \)). The frequency of APOE*4 carriers (39% versus 21%) and APOE*4 allele (21% versus 11%) was almost double in cases than in controls (Table 2). The odds ratios (ORs) with one and two copies of APOE*4 were 2.22 (95% CI: 1.80–2.73) and 4.92 (95% CI: 3.99–6.06), respectively. On the other hand, the frequency of the APOE*2 allele was similar in AD Dementia and non-dementia groups (6.94% versus 8.15%) with a non-significant OR of 0.863 (95% CI: 0.73–1.01; \( p = 0.3245 \)). Similar results were obtained in the Any Dementia analysis (Supplementary Table 1).

**Association analysis of known AD-loci**

Of the reported AD-associated top SNPs in 93 loci among Whites [5], 76 SNPs passed QC in the GEM sample (Supplementary Table 2). In addition
to APOE, ten reported top SNPs achieved nominal significance in the in the GEM sample in the following loci: OR2B2/rs1497525 (p = 5.16E-03), HS3ST5/ rs785129 (p = 1.64E-02), GPR41/rs2718058 (p = 4.14E-04), CTSB/rs1065712 (p = 5.53E-02), CC DC6/rs1171814 (p = 1.80E-03), PICALM/rs867611 (p = 1.85E-02), MSA6A/rs7933202 (p = 4.25E-02), APH1B/rs117618017 (p = 1.90E-02), PLCG2/rs12444183 (p = 4.76E-02), and ABCA7/rs3752246 (p = 1.49E-02). Although top SNPs at an additional 47 loci showed the same directional effects in the GEM sample as the ones reported in very large data sets, they did not achieve statistically significance, perhaps due to the relatively small sample size of the dementia group in this study.

Genome-wide association analyses

The quantile-quantile plots in the AD and Any Dementia analyses showed that the observed p-values largely adhered to the p-values of the null distribution until the deviation at the tail (Fig. 1a; Supplementary Figure 1). This indicated some associations and a low likelihood of false positives results.

Two genome-wide significant (GWS) signals were observed in the AD Dementia analysis, including the known association on chromosome 19 and a novel signal on chromosome 12 (Fig. 1b). Similar results were observed in the Any Dementia analysis (Supplementary Figure 2; Supplementary Table 3). There were six SNPs in the APOE region that reached the GWS threshold of p < 5E-08 (Table 3). These associations are due to the strong association of APOE*4/rs429358 (OR = 2.22; p = 9.36E-14) with AD and its LD with the surrounding SNPs (Fig. 2).

The novel signal observed on chromosome 12 was characterized by three GWS SNPs: rs148377161 (p = 1.66E-08), rs148760255 (p = 2.09E-08), and rs192213585 (p = 2.64E-08) (Table 3). While two of these SNPs were imputed (rs148377161 and rs192213585), rs148760255 was genotyped. The presence of a genotyped SNP in this region reduces the likelihood of this association being caused by an error in imputation. All three SNPs are rare (MAF: 0.019 - 0.021) and were in LD with each other (pairwise r² = 0.73, 0.80, 0.83) (Fig. 3). The regional association plot shows that these three SNPs are located near the CCDC91 (coiled-coil domain containing 91) gene such that rs148760255 and s148377161 are located 53 kb and 127 kb, respectively downstream from CCDC91 (Fig. 4a). Based on the NCBI's gene prediction program Gnomon and GRCh38 assembly (ncbi.nlm.nih.gov/), there are also two predicted non-coding RNAs (ncRNAs) genes, LOC105369711 and sni LOC101928705, near CCDC91. While rs192213585 is located within LOC105369711, rs148377161 is 61 kb downstream from LOC105369711 and rs148760255 is 24 kb downstream from LOC101928705 (Fig. 4b).

In order to check replication of the chromosome 12 signal, we examined these SNPs in the International Genomics of Alzheimer’s Project (IGAP) discovery sample of 63,926 subjects [13]. Although the association directions of these SNPs in IGAP were the same, they were not statistically significant (p = 0.862 for rs148377161; p = 0.166 for rs192213585; and p = 0.248 for rs148760255). Because of the strong LD between these SNPs, we expected to observe similar p-values for the three SNPs in the IGAP sample, as we had observed in our sample, but this was not the case. This may be due to inconsistent quality of the imputation in this region obtained in the different sets of samples used in IGAP.

### Table 2

<table>
<thead>
<tr>
<th>Genotype</th>
<th>AD Dementia (n = 382)</th>
<th>Cognitively Normal (n = 2,197)</th>
<th>Total (n = 2,579)</th>
</tr>
</thead>
<tbody>
<tr>
<td>E2/2</td>
<td>3 (0.79%)</td>
<td>18 (0.82%)</td>
<td>21 (0.81%)</td>
</tr>
<tr>
<td>E2/3</td>
<td>30 (7.85%)</td>
<td>279 (12.70%)</td>
<td>309 (11.98%)</td>
</tr>
<tr>
<td>E2/4</td>
<td>17 (4.45%)</td>
<td>43 (1.96%)</td>
<td>60 (2.33%)</td>
</tr>
<tr>
<td>E3/3</td>
<td>199 (52.09%)</td>
<td>1,442 (65.64%)</td>
<td>1,641 (63.63%)</td>
</tr>
<tr>
<td>E3/4</td>
<td>124 (32.46%)</td>
<td>391 (17.80%)</td>
<td>515 (19.97%)</td>
</tr>
<tr>
<td>E4/4</td>
<td>9 (2.36%)</td>
<td>24 (1.09%)</td>
<td>33 (1.28%)</td>
</tr>
<tr>
<td>Hardy-Weinberg p</td>
<td>0.079</td>
<td>0.878</td>
<td>0.402</td>
</tr>
<tr>
<td>Allele Frequency</td>
<td>0.0694</td>
<td>0.0815</td>
<td>0.0797</td>
</tr>
<tr>
<td>APOE*2</td>
<td>0.7225</td>
<td>0.8089</td>
<td>0.7961</td>
</tr>
<tr>
<td>APOE*4</td>
<td>0.2082</td>
<td>0.1097</td>
<td>0.1243</td>
</tr>
</tbody>
</table>
Fig. 1. a) Quantile-quantile plot showing observed versus expected $p$-values for the AD Dementia analysis. b) Manhattan plot showing genome-wide $p$-values for the association with AD. The threshold for genome-wide significance ($p < 5E-08$) is indicated by the red line and the threshold for suggestive significance ($p < 1E-05$) is indicated by the blue line.

Table 3

<table>
<thead>
<tr>
<th>CHR</th>
<th>SNP</th>
<th>BP</th>
<th>A1</th>
<th>MAF Cases</th>
<th>MAF Controls</th>
<th>OR (95% CI)</th>
<th>$p$</th>
<th>LOC</th>
<th>GENE/LOCUS</th>
</tr>
</thead>
<tbody>
<tr>
<td>19</td>
<td>rs429358</td>
<td>45411941 C</td>
<td>0.2060</td>
<td>0.1099</td>
<td>2.217 (1.798–2.733)</td>
<td>9.356E-14</td>
<td>Exonic</td>
<td>APOE</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>rs4420638</td>
<td>45422946 G</td>
<td>0.2267</td>
<td>0.1437</td>
<td>1.791 (1.474–2.176)</td>
<td>4.477E-09</td>
<td>Downstream</td>
<td>APOC1</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>rs12721051</td>
<td>45422160 G</td>
<td>0.2254</td>
<td>0.1435</td>
<td>1.785 (1.469–2.169)</td>
<td>5.575E-09</td>
<td>Intronic</td>
<td>APOC1</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>rs769449</td>
<td>45410002 A</td>
<td>0.1606</td>
<td>0.0891</td>
<td>2.027 (1.613–2.546)</td>
<td>1.278E-09</td>
<td>Intronic</td>
<td>APOE</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>rs6857</td>
<td>45392254 T</td>
<td>0.2189</td>
<td>0.1392</td>
<td>1.784 (1.463–2.175)</td>
<td>1.052E-08</td>
<td>3' UTR</td>
<td>NECTIN2</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>rs7256200</td>
<td>45415935 T</td>
<td>0.1580</td>
<td>0.0913</td>
<td>1.913 (1.526–2.399)</td>
<td>1.924E-08</td>
<td>Upstream</td>
<td>APOC1</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>rs192213585</td>
<td>28918097 C</td>
<td>0.0479</td>
<td>0.0168</td>
<td>3.245 (2.144–4.913)</td>
<td>2.644E-08</td>
<td>Intergenic</td>
<td>CCDC91/LOC105369711</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>rs148760255</td>
<td>28755421 G</td>
<td>0.0453</td>
<td>0.0152</td>
<td>3.408 (2.220–5.233)</td>
<td>2.087E-08</td>
<td>Intergenic</td>
<td>CCDC91/LOC105369711</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>rs148377161</td>
<td>28830097 T</td>
<td>0.0479</td>
<td>0.0163</td>
<td>3.312 (2.185–5.019)</td>
<td>1.661E-08</td>
<td>Intergenic</td>
<td>CCDC91/LOC105369711</td>
<td></td>
</tr>
</tbody>
</table>

CHR, chromosome; SNP, single-nucleotide polymorphism; BP, base-pair position based on GRCh37 assembly; A1, effect allele; MAF, minor allele frequency of the effect allele; OR, odds ratio with 95% confidence interval; LOC, location of the SNP.
Fig. 2. LD heatmap for chromosome 19 signal in the APOE region.

Fig. 3. LD heatmap for chromosome 12 signal in the CCD91/LOC105369711 region.
**Sex-stratified analysis**

Sex-stratified quantile-quantile plots in AD Dementia and Any Dementia are shown in Supplementary Figures 3–5 and 7, and the Manhattan plot in males in Supplementary Figure 8. The sex-stratified analysis showed GWS at the APOE locus only in females in both the AD Dementia and Any Dementia groups (Fig. 5; Supplementary Figure 6; Table 4; Supplementary Table 4). However, among males, we observed a GWS signal in the AD Dementia analysis on chromosome 1 with rs140076909 (OR = 4.51, 95% CI: 2.64–7.70; \( p = 3.72 \times 10^{-8} \)) (Fig. 6a; Table 4). Although this SNP did not achieve the GWS threshold in Any Dementia males, the strength of association was similar with an OR of 4.13 (95% CI: 2.42–7.03; \( p = 1.94 \times 10^{-7} \)) (Supplementary Table 4).

The top SNP, rs140076909, is located 70 kb downstream of ncRNA LOC729987 and 319 kb upstream from SNX7 (Sorting Nexin 7) (Fig. 6a). Another SNP, rs117803234, located on chromosome 7 almost reached GWS in males (\( p = 6.93 \times 10^{-8} \); Fig. 6b; Table 4), which is located 18 kb upstream from a pseudogene LOC100131257 and 66 kb upstream from C1GALT1 (core 1 synthase, glycoprotein-N-acetylgalactosamine 3-beta-galactosyltransferase 1). Based on the NCBI’s gene prediction program Gnomon and GRCh38 assembly, rs117803234 is 11 kb downstream of a predicted ncRNA LOC105375138 as well as a cis-regulatory element, LOC116183083 (Fig. 6c). These SNPs were also examined in the available sex-combined data from the IGAP GWAS, but neither rs140076909 (\( p = 0.823 \)) nor rs117803234 (\( p = 0.695 \)) was significant.
Fig. 5. a) Manhattan plot showing genome-wide $p$-values for the association of Alzheimer’s Dementia in females. The threshold for genome-wide significance ($p < 5 \times 10^{-8}$) is indicated by the red line and the threshold for suggestive significance ($p < 1 \times 10^{-5}$) is indicated by the blue line. 

b) Manhattan plot showing genome-wide $p$-values for the association of Alzheimer’s Dementia in males. The threshold for genome-wide significance ($p < 5 \times 10^{-8}$) is indicated by the red line and the threshold for suggestive significance ($p < 1 \times 10^{-5}$) is indicated by the blue line.

equivalent sex-combined $p$-values in our study were: rs140076909 ($p = 1.93 \times 10^{-4}$ in AD Dementia; $p = 8.99 \times 10^{-4}$ in Any Dementia), rs117803234 ($p = 4.69 \times 10^{-3}$ in AD Dementia; $p = 6.05 \times 10^{-3}$ in Any Dementia).

**APOE*4 and non-APOE*4-stratified analyses**

The results of APOE*4 stratified analyses are presented in Supplementary Tables 5–8 and Supplementary Figures 9–14. Although there were no GWS signals, multiple suggestive associations ($p < 1 \times 10^{-5}$) were observed in both APOE*4 and non-APOE*4 carriers. There were three most significant associations among non-APOE*4 carriers in AD Dementia: HSP90AA6P/rs111618415, $p = 1.89 \times 10^{-7}$ on chromosome 4; RPL29P13/rs187039392, $p = 8.98 \times 10^{-7}$ on chromosome 5 and BORCS5/rs556343211, $p = 8.61 \times 10^{-7}$ on chromosome 12 (Supplementary Table 8). Likewise, two most significant associations were observed among non-APOE*4 carriers in Any Dementia: IRF2BP2/rs35581844, $p = 2.22 \times 10^{-7}$ on
Table 4  

<table>
<thead>
<tr>
<th>Chr</th>
<th>SNP</th>
<th>BP</th>
<th>MAF Cases</th>
<th>MAF Controls</th>
<th>OR (95% CI)</th>
<th>p</th>
<th>LOC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>rs140076909</td>
<td>9880825</td>
<td>0.0614</td>
<td>0.0150</td>
<td>4.506 (2.636–7.702)</td>
<td>3.72E-08</td>
<td>Intergenic</td>
</tr>
<tr>
<td>7</td>
<td>rs117803234</td>
<td>7155404</td>
<td>0.0711</td>
<td>0.0195</td>
<td>3.765 (2.325–6.096)</td>
<td>6.93E-08</td>
<td>Intergenic</td>
</tr>
<tr>
<td>12</td>
<td>rs148760255</td>
<td>28755421</td>
<td>0.0417</td>
<td>0.0179</td>
<td>2.578 (1.436–4.631)</td>
<td>1.52E-03</td>
<td>Intergenic</td>
</tr>
<tr>
<td>12</td>
<td>rs148377161</td>
<td>28830097</td>
<td>0.0417</td>
<td>0.0183</td>
<td>2.500 (1.395–4.482)</td>
<td>2.09E-03</td>
<td>Intergenic</td>
</tr>
<tr>
<td>12</td>
<td>rs192213585</td>
<td>28918097</td>
<td>0.0441</td>
<td>0.0183</td>
<td>2.679 (1.510–4.753)</td>
<td>7.56E-04</td>
<td>Intergenic</td>
</tr>
<tr>
<td>19</td>
<td>rs429358</td>
<td>45410082</td>
<td>0.1789</td>
<td>0.1075</td>
<td>1.875 (1.400–2.451)</td>
<td>2.43E-05</td>
<td>Exonic</td>
</tr>
</tbody>
</table>

**Bioinformatics analyses**

Next, the significant SNPs on chromosome 12 locus as well as the sex-stratified significant SNPs on chromosome 1 and 7 locus were examined to determine their possible functional significance (Supplementary Table 9). Using data from the Genotype-Tissue Expression (GTEx) v8 database these SNPs were investigated for any single-tissue eQTLs. None of these SNPs from these three loci had any significant eQTLs in any tissue. These SNPs were also investigated in the RegulomeDB database where a lower number rank on a scale of 1–7 indicates higher evidence for regulatory function and a higher probability score on a scale of 1–2 indicates a higher probability of regulatory function. CADD scores were obtained for these five SNPs where a higher score indicates a higher level of deleteriousness. The highest CADD score of 4.32 was observed for the top chromosome 1 SNP, rs140076909. We also examined four SNPs in LD with the top SNPs on chromosomes 1 and 12. One SNP on chromosome 1 (rs184006746) in LD with the top SNP (R^2 = 0.76) had a RegulomeDB score of 3a, indicating a 65% probability of being regulatory function. This SNP also showed the second highest CADD score of 3.49 (Supplementary Table 9).

**DISCUSSION**

In this community-based sample of older White subjects, we have confirmed the established association of APOE^*4 with AD dementia where ORs with one and two copies of APOE^*4 were 2.22 and 4.92, respectively. This observation is similar to the majority of reported associations in community-based White populations [14], but weaker than the reported ORs of about 3.5 with one copy of APOE^*4 and 14.5 with two copies of APOE^*4 in standard case-control studies among Whites [15, 16] where AD cases are largely derived from AD centers having moderate to severe AD compared to the mild dementias in community-based cohorts. We also found that the association of APOE^*4 is stronger in females (OR = 2.72; p = 1.74E-10) than in males (OR = 1.88; p = 2.43E-05) in the GEM sample. A previous study...
Fig. 6. a) Regional association plot on chromosome 1 centered around the most significant SNP, rs140076909 ($p = 3.72 \times 10^{-8}$) based on GRCh37 assembly. b) Regional association plot on chromosome 7 centered around the most significant SNP, rs117803234 ($p = 6.93 \times 10^{-8}$) based on GRCh37 assembly. c) Locations of predicted genes and top SNP, rs117803234 relative to LOC100131257, LOC105375138, and LOC116183083 based on the NCBI’s gene prediction program Gnomon and GRCh38 assembly.
found a similar sex difference in APOE*4 association, though the effect size was not statistically significant between females (OR = 3.66; p = 1.78E-215) and males (OR = 3.25; p = 5.04E-122) [17]. Of the other known AD loci, reported SNPs near or at OR2B2, HS3ST5, GPR141, CTSD, CCDC6, MS4A6A/MS4A4E, PICALM, APH1B, PLCG2, and ABCA7 were also replicated in this study at nominal significance.

We have identified a novel GWS locus on chromosome 12 in this community-based sample of older subjects. However, this finding was not replicated in the IGAP-2019 data. The lack of replication may be due to a false-positive association observed in this small sample of AD subjects. Alternatively, this is a real association in a unique older sample with incident dementia that needs to be confirmed in an equivalent older and community-based sample. The three GWS SNPs on chromosome 12 are located near CCDC91. These SNPs lacked evidence of being significant eQTLs and had relatively low CADD scores. The RegulomeDB ranks of two of these SNPs, however, were 5, indicating a transcription factor binding location or a DNase peak, which is correlated with transcriptional activity. Most importantly, there are two predicted ncRNA genes, LOC105369711 and LOC101928705, near CCDC91 where one of the three top SNPs (rs192213585) is located within LOC105369711, and the other two SNPs are only 24–61 kb downstream from these ncRNA genes. In the sex-stratified analysis, we also identified two novel loci near ncRNA LOC729987 on chromosome 1 and ncRNA LOC105375138 on chromosome 7 in males. ncRNAs are relatively new classes of RNA that are broadly divided into small ncRNAs (<200 nucleotides), including microRNAs and piwi-interacting RNAs, and long ncRNAs (>200 nucleotides) including circular and linear RNAs. These ncRNAs have the ability to regulate gene expression and affect metabolic pathways and have been implicated in neurodegenerative diseases [18–21].

In conclusion, our genome-wide analyses on a unique community-based sample of older subjects have identified three novel loci associated with incident dementia, in addition to APOE. The novel loci are located near ncRNAs genes with potential regulatory effects. Further genetic and functional characterization of these regions would help to elucidate these associations with AD. These novel associations should be considered provisional until replicated in larger and older incident dementia cases derived from community-based studies. Our study suggests that the use of similar samples of older adults who have developed incident dementia over time may be a helpful tool in unraveling novel genetic associations normally not seen in highly selected cases derived from AD centers or similar referral centers.

ACKNOWLEDGMENTS

This study was supported by the National Institute on Aging (NIA) grants AG030653, AG041718, and AG064877. A subset of samples used in this study were obtained from the National Centralized Repository for Alzheimer’s Disease and Related Dementia (NCRAD), which receives government support under a cooperative agreement grant (U24 AG021886) awarded by the NIA. We thank contributors who collected samples used in this study, as well as patients and their families, whose help and participation made
this work possible. This publication was made possible by Grant Number U01 AT000162 from the National Center for Complementary and Alternative Medicine, National Institutes of Health.

Authors’ disclosures available online (https://www.j-alz.com/manuscript-disclosures/22-0293r1).

SUPPLEMENTARY MATERIAL

The supplementary material is available in the electronic version of this article: https://dx.doi.org/10.3233/JAD-2202923.

REFERENCES

[10] Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Muller J, Sklar P, de Bakker PI, Daly MJ, Sham PC (2007) PLINK: A toolset for whole-genome association and population-based linkage analysis. Am J Hum Genet 81, 559-575.