APOE-stratified genome-wide association study suggests potential novel genes for late-onset Alzheimer's disease in East-Asian descent

Sarang Kang\textsuperscript{a;b,1}, Jungsoo Gim\textsuperscript{a;b;c;1,*}, Tamil Iniyan Gunasekaran\textsuperscript{a;c}, Kyu Yeong Choi\textsuperscript{a}, Jang Jae Lee\textsuperscript{a}, Eun Hyun Seo\textsuperscript{a;d}, Pan-Woo Ko\textsuperscript{e}, Ji Yeon Chung\textsuperscript{a;f}, Seong-Min Choi\textsuperscript{a;g}, Young Min Lee\textsuperscript{h}, Jee Hyang Jeong\textsuperscript{i}, Kyung Won Park\textsuperscript{j}, Min Kyung Song\textsuperscript{a;k}, Ho-Won Lee\textsuperscript{e}, Ki Woong Kim\textsuperscript{l;m}, Seong Hye Choi\textsuperscript{n}, Dong Young Lee\textsuperscript{o}, Sang Yun Kim\textsuperscript{p}, Hoowon Kim\textsuperscript{a;f}, Byeong C. Kim\textsuperscript{a;g}, Takeshi Ikeuchi\textsuperscript{q}, and Kun Ho Lee\textsuperscript{a;b;c;r;s,*}

\textsuperscript{a}Gwangju Alzheimer's & Related Dementias Cohort Research Center, Chosun University, Gwangju 61452, Republic of Korea
\textsuperscript{b}Department of Integrative Biological Sciences and BK21 FOUR Educational Research Group for Age-Associated Disorder Control Technology, Chosun University, Gwangju 61452, Republic of Korea
\textsuperscript{c}Department of Biomedical Science, Chosun University, Gwangju 61452, Republic of Korea
\textsuperscript{d}Department of Premedical Science, Chosun University College of Medicine, Gwangju 61452, Republic of Korea
\textsuperscript{e}Department of Neurology, Kyungpook National University School of Medicine, Daegu 41566, Republic of Korea
\textsuperscript{f}Department of Neurology, Chosun University Hospital, Gwangju 61452, Republic of Korea
\textsuperscript{g}Department of Neurology, Chonnam National University Medical School, Gwangju 61469, Republic of Korea
\textsuperscript{h}Department of Psychiatry, Pusan National University School of Medicine, Busan 50612, Republic of Korea
\textsuperscript{i}Department of Neurology, Ewha Womans University Mokdong Hospital, Ewha Womans University School of Medicine, Seoul 07985, Republic of Korea

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Department of Neurology, Donga University College of Medicine, Busan 49315, Republic of Korea

Chonnam National University Gwangju 2nd Geriatric Hospital, Gwangju 61469, Republic of Korea

Department of Psychiatry and Behavioral Science, Seoul National University College of Medicine, Seoul 03080, Republic of Korea

Department of Brain and Cognitive Science, Seoul National University College of Natural Sciences, Seoul 03080, Republic of Korea

Department of Neurology, Inha University School of Medicine, Incheon 22212, Republic of Korea

Department of Neuropsychiatry, Seoul National University Hospital, Seoul 03080, Republic of Korea

Department of Neurology, Seoul National University Bundang Hospital, Seongnam 13620, Republic of Korea

Department of Molecular Genetics, Brain Research Institute, Niigata University, Niigata 950-2181, Japan

Korea Brain Research Institute, Daegu 41062, Republic of Korea

Neurozen Inc., Seoul 06236, Republic of Korea

Note: [1] These authors contributed equally to this work.

Correspondence: [*] Correspondence to: Jungsoo Gim, PhD, Department of Biomedical Science, Chosun University, Gwangju 61452, Republic of Korea, Email: jgim@chosun.ac.kr and Kun Ho Lee, PhD, Department of Biomedical Science, Chosun University, Gwangju 61452, Republic of Korea, Email: leekho@chosun.ac.kr

Running title: East Asian-specific novel loci associated with LOAD
ABSTRACT

In this study, we report two new possible susceptible genes for late-onset Alzheimer’s disease identified through an APOE-stratified genome-wide association analysis (GWAS) using East Asian samples. In the discovery phase, we performed a GWAS of Alzheimer’s disease in 2,291 Korean seniors from the Gwangju Alzheimer’s and Related Dementias (GARD) cohort study. A successive replication analysis with a Japanese sample of size 1,956 suggested three novel susceptible SNPs in two genes: LRIG1 and CACNA1A. This study demonstrates that the discovery of AD-associated variants can be accomplished in non-European ethnic groups with a more homogeneous genetic background using samples comprising fewer subjects.

Keywords: Genome-wide association study (GWAS), Alzheimer’s disease, Late-onset Alzheimer’s disease, APOE, Stratified genome analysis

INTRODUCTION

A genetic component has consistently been identified as an important antecedent factor in Alzheimer’s disease (AD), and Apolipoprotein E (APOE) is the most prevalent risk factor of the disease [1, 2]. The effect of the APOE genotype, however, varies greatly among populations [3]. One noteworthy example is that the e4 allele, the risk form of APOE, is virtually absent among Arabs living in the northern Israeli community where the prevalence of dementia (in spite of unspecified pathology) is roughly double that in European ancestry [4].

Since the identification of the only locus in APOE with a sample consisting of 1,086
individuals in 2007 [5], genome-wide association studies (GWASs) have increased the sample size to improve the statistical power to identify the missing causal variants for late-onset AD (LOAD) [6, 7]. The latest and largest GWAS expands to include 1,126,563 cases and controls (including 364,859 proxy) [Douglas P. Wightman, et al., 2020, medRxiv: https://doi.org/10.1101/2020.11.20.20235275]. The variants and associated genes from those GWASs were identified mostly in individuals of European ancestry and found enriched in various biological pathways, including immune responses and endocytosis, suggesting genetic heterogeneity or the complex nature of the LOAD.

Genetic studies in diverse populations have increased our understanding of the genetic architecture of LOAD [8-12]. Studies of non-European ancestry populations might have a chance of being used to discover variants that are rare or absent, thus showing a smaller effect size in European ancestry [13]. Notably, the discovery of such AD-susceptible loci, displaying allelic heterogeneity among diverse populations, has been accomplished with non-European ancestry groups of a more homogeneous genetic background using samples comprising several thousand or much fewer subjects [10, 11].

In the present study, we leverage the genetic architecture of Koreans, a homogeneous East Asian population maintaining a distinct genetic profile with a high prevalence of AD among seniors (65 years and older), to discover AD-associated variants. For this purpose, we first performed GWAS and post-hoc GWAS with APOE stratified in Koreans using the Korea Biobank Array (referred to as Korean Chip) designed for Koreans [14] and replicated the findings by combining genotype level data from
another East Asian population [11].

MATERIALS AND METHODS

Study Sample

We analyzed genotype level data of 2,291 individuals (1,119 AD cases and 1,172 controls) for the discovery stage, and 1,956 individuals (980 AD cases and 976 controls) for the replication stage (Table 1). Participants for the discovery stage were sampled from the Gwangju Alzheimer’s & Related Dementias (GARD) Cohort Research Center, a longitudinal single-center study designed to develop clinical, imaging, genetic, and other biomarkers for early detection and tracking of dementia (mostly AD) from local senior citizens aged 60 years or older in Gwangju, South Korea. The study protocol was approved by the Institutional Review Board of Chosun University Hospital, Korea (CHOSUN 2013-12-018-070). All volunteers or authorized guardians for cognitively impaired individuals gave written informed consent before participation. Out of 16,002 participants in the GARD cohort, 2,291 (1,119 AD dementia and 1,172 cognitive normal) were chosen to maximize contrast between AD cases and cognitive normal controls for the GWAS. All cases were at least 60 years old and met NINCDS-ADRDA criteria for AD [15], while cognitively healthy controls were at least 70 years old and either judged to be cognitively normal or did not meet pathological criteria (Supplementary Material).

Genotyping, Imputation, and Gene Mapping

We applied extensive quality control measures used in the standard protocol [16] and
excluded subjects with a low genotyping success rate (95%), heterozygosity outliers and those that exhibited cryptic relatedness. We assessed study population heterogeneity by means of a principal component analysis (PCA) and multidimensional scaling (MDS) and removed those that did not meet the criteria before proceeding with further analyses. We also excluded those with SNPs with a minor allele frequency <1%, a genotyping success rate <95%, or a deviation of the genotype distribution from Hardy-Weinberg equilibrium in the control group (P < 1×10⁻⁶). To increase the genotyping coverage between the two datasets, we imputed missing genotypes and SNPs using pre-phased reference haplotypes from the Haplotype Reference Consortium (HRC) panel version 1.1 for each dataset. We subjected all imputed SNPs to the same quality criteria described above and added the requirement of imputation quality with an info score >0.5, leaving 35,685,761 and 39,044,005 SNPs for discovery and replication analysis, respectively. Functional mapping of SNPs to genes was conducted using a gene annotation file from Affymetrix [14] and the SNPnexus web tool [17].

**Statistical analysis**

A set of GWAS analyses for AD status was performed in three models of a two-stage framework (discovery and replication): 1) all samples; 2) APOE e4 carriers; and 3) APOE e4 non-carriers. The GWAS was performed separately in each model for both the discovery and replication stages. The association of genotype dosages of each SNP in the additive component with the AD case-control status was estimated using logistic regression analysis adjusting for age, sex, and first four principal components using PLINK and EIGENSTRAT software [18-20]. Only the SNPs that attained a
suggestively significant association level ($P < 5 \times 10^{-5}$) were tested in the replication stage using a Japanese dataset. A total of 4,247 subjects was assembled for meta-analysis and SNP effect estimates and their SEs were combined by a fixed effect model with inverse variance weighted method using the METAL software [21]. SNP heritability was estimated using the GCTA software [22].

RESULTS

**Genome-wide inference for AD status**

For discovery, we first conducted GWA analysis with the GARD sample, which included 2,291 subjects and identified 54 genome-wide significant ($5 \times 10^{-8}$) SNPs across three genes (genomic inflation = 1.03), and all those genes ($APOE$, $PVRL2$, and $TOMM40$) were found to be associated with LOAD in previous studies (Table 2). We estimated the liability-scale SNP heritability, excluding those loci, and a relatively large estimate ($0.566 \pm 0.152$) was obtained. In addition, a portion of SNPs in other previously known genes, such as $ABCA7$ and $BIN1$, was replicated at a genome-wide suggestive level ($5 \times 10^{-5}$). These results suggest that additional common variants with smaller effects (or possibly rarer variants with larger effects) yet to be found might be identified through a post-hoc analysis with a less stringent statistical significance.

We carried out $APOE$-stratified GWA analysis using subgroups of $APOE$ e4 carrying status. A total of 219 variants (annotated in 61 genes) and 306 variants (in 82 genes) were respectively identified as e4 carriers (genomic inflation = 1.02) and non-carriers (genomic inflation = 1.03) with a suggestive level of $5 \times 10^{-5}$. Loosening statistical
significance might increase false findings; thus, we included an independent cohort dataset for replication. For valid novel discovery, we performed a meta-analysis with the replication dataset, filtered with previous GWAS signals, resulting in examination of 7 out of 61 and 15 out of 82 genes for e4 carriers and non-carriers, respectively. Considering the effect direction and statistical significance of the variants in both the discovery and replication datasets, CACNA1A (OR = 1.726, lead SNP rs189753894) and LRIG1 (OR = 0.539, lead SNP rs2280575) remained in the association result for further functional annotation (Table 2).

**Functional annotation**

We followed up the association results of two novel genes, CACNA1A and LRIG1, by carrying out functional annotation using web-based platforms from the Human Protein Atlas (HPA) [23], Genotype-Tissue Expression (GTEx) [24], and Functional Annotation of Mammalian Genomes 5 (FANTOM5) [25]. CACNA1A, a risk gene candidate, is on chromosome 19. However, it is not likely to be influenced by the APOE signal because CACNA1A is far from the APOE locus (more than 32Mb) and is identified with APOE non-carriers. We also confirmed its independence with all subjects by conditional analysis with APOE*4 (rs429358). Tissue specificity analysis with GTEx, FANTOM5, and HPA showed that CACNA1A is expressed in brain tissues, particularly in the cerebellum at both the transcription and protein levels (Supplementary Fig. 1A, Fig. 1A). The CACNA1A gene plays an important role in neuronal cell death. It codes subunits of neuronal calcium channels, which are involved in the neuronal cell death and toxicity of amyloid beta [26, 27].

Expression levels of LRIG1 were high in the cerebral cortex at 74.1 protein-coding
transcripts per million (pTPM) in HPA, 16.0 pTPM reported by GTEx (Supplementary Fig. 1B). In the FANTOM5 dataset, the RNA expression levels of LRIG1 were higher in the brain tissues than in other tissues, especially for the hippocampal formation, which was 179.9 scaled tags per million. eQTL analysis with lead SNP rs2280575 in LRIG1 showed significant results in the hypothalamus (Fig. 1B). The association between LRIG1 and the hippocampus was evaluated in in vitro and in vivo analyses in a previous study [28].

DISCUSSION

Here, we conducted GWA analysis with a Korean sample to identify AD-associated loci. We validated the association between AD status and APOE locus in chromosome 19 with genome-wide significance. To mask the effect of APOE and identify novel candidate loci in East Asians, we carried out an APOE4 stratified GWAS with two independent East Asian samples and identified CACAN1A (lead SNP rs189753894) and LRIG1 (lead SNP rs2280575) among APOE4 non-carriers.

From the functional annotation and literature survey, we found that the two genes play an essential role in neurobiological function. CACNA1A, a subunit of the neuronal calcium channel, is predominantly and specifically expressed in brain tissue and was found to be involved in neuronal cell death and toxicity of amyloid beta, suggesting a plausible risk gene for AD. The genetic function of LRIG1 has been shown to be involved in the dendritic formation of neurons, as well as in affecting the neural cell function of the hippocampus. Abnormal expression and dysfunction of LRIG1 lead to dendritic abnormalities, which are involved in morphogenesis of hippocampal
dendrites by brain-derived neurotrophic factor (BDNF) signaling. BDNF plays a major role in the growth, development and survival of neurons, and is involved in synaptic plasticity and synaptogenesis for learning and memory in the adult brain [29-31]. BDNF showed low gene expression levels in patients with AD [32]. Several studies have shown that high expression levels of BDNF could slow down cognitive decline in the elderly and in patients with AD [29, 33-35].

There are limitations to this study that need to be considered. The sample size in this study is relatively smaller than those in other AD GWA studies. However, the sample used in this analysis was a representative sample with well-defined cases and controls: subjects to maximize contrast were chosen, and any subjects with ambiguous diagnosis were removed from the further analysis. Although the previous studies about the function of CACNA1A and LRIG1 suggest their role in the development of the LOAD, a conclusion whether each gene has an impact on the progression of dementia or protecting neurons from degeneration requires further biological validation with an AD model system.

To conclude from the findings made in this work, we validated a portion of previously reported LOAD-associated genes, identified from the mostly Caucasian subjects, using the Koreans. We also identified and replicated East Asian-specific novel loci in CACNA1A and LRIG1 through a post-hoc GWAS with APOE4 stratified. The finding in this work might provide an improved understanding of complex genetic signature of LOAD.
Acknowledgments  This research was supported by the Original Technology Research Program for Brain Science of the National Research Foundation (NRF-2014M3C7A1046041 and NRF-2014M3C7A1046042), and KBRI basic research program (20-BR-03-02) funded by Koera Ministry of Science and ICT. This study was also supported by Japan Agency for Medical Research and Development (AMED) (JP18kk0205009).

IRB  The study protocol was approved by the Institutional Review Board of Chosun University Hospital, Korea (CHOSUN 2013-12-018-070). All volunteers or authorized guardians for cognitively impaired individuals gave written informed consent before participation.

Conflict of Interest/Disclosure Statement  The authors have no conflict of interest to report
REFERENCES


[30] Lu B, Nagappan G, Lu Y (2014) BDNF and synaptic plasticity, cognitive function, and...


FIGURE LEGENDS AND TABLES

Table 1. Summary of demographic information of each group in GWAS analyses.

Table 2. Top-ranked genome-wide association results in the Korean discovery sample (P < 5x10^-5) and their replication in Japanese and meta-analyses.

Figure 1. Gene expression of CACNA1A across different tissues (B) and expression quantitative trait loci (eQTL) analysis of rs2280575 in LRIG1 (A). (A) The CACNA1A is expressed at a high level in the cerebellum (median TPM: 304.8) and in the cerebellar hemisphere (median TPM: 275.0). (B) Violin plot showing the effect of the eQTL rs2280575 on LRIG1 expression in hypothalamus (P = 1.13×10^-2). (Image available from https://gtexportal.org/home/gene/CACNA1A).
### Table 1. Summary of demographic information for each group in GWAS analyses

<table>
<thead>
<tr>
<th></th>
<th>Discovery (n=2,291)</th>
<th>Replication (n=1,956)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CN</td>
<td>AD</td>
</tr>
<tr>
<td>All samples, n</td>
<td>1,172</td>
<td>1,119</td>
</tr>
<tr>
<td>Female, n (%)</td>
<td>661 (56.4)</td>
<td>715 (63.9)</td>
</tr>
<tr>
<td>Age, m (s.d)</td>
<td>76.03 (8.9)</td>
<td>74.60 (8.9)</td>
</tr>
<tr>
<td>APOE e4 non-carrier, n</td>
<td>976</td>
<td>621</td>
</tr>
<tr>
<td>Female, n (%)</td>
<td>544 (55.7)</td>
<td>393 (63.3)</td>
</tr>
<tr>
<td>Age, m (s.d)</td>
<td>76.10 (8.9)</td>
<td>75.43 (9.1)</td>
</tr>
<tr>
<td>APOE e4 carrier</td>
<td>196</td>
<td>498</td>
</tr>
<tr>
<td>Female, n (%)</td>
<td>117 (59.7)</td>
<td>322 (64.7)</td>
</tr>
<tr>
<td>Age, m (s.d)</td>
<td>95.45 (8.9)</td>
<td>73.32 (9.0)</td>
</tr>
</tbody>
</table>
Table 2. Top-ranked GWAS result in the discovery, replication and meta-analysis.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Chr</th>
<th>Lead SNP (type)*</th>
<th>BP</th>
<th>m/M</th>
<th>MAF</th>
<th>Discovery (1,172/1,119)</th>
<th>Replication (976/980)</th>
<th>Meta-analysis</th>
<th>Previously reported GWAS</th>
<th>Alt Allele Frq†</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>OR (95% CI)</td>
<td>OR (95% CI)</td>
<td>P</td>
<td>OR (95% CI)</td>
<td>OR (95% CI)</td>
<td>P</td>
<td>OR</td>
<td>P</td>
<td>GENE (SNP)</td>
</tr>
<tr>
<td>ALL samples</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>APOE</td>
<td>19</td>
<td>rs429358 (m)</td>
<td>3.637 (3.03-4.37)</td>
<td>0.169</td>
<td>3.74×10⁻⁴³</td>
<td>2.608 (2.04-3.33)</td>
<td>1.89×10⁻¹⁴</td>
<td>3.227</td>
<td>5.90×10⁻⁵⁵</td>
<td>rs429358</td>
</tr>
<tr>
<td>TOMM40</td>
<td>19</td>
<td>rs10119 (3')</td>
<td>3.378 (2.83-4.03)</td>
<td>0.180</td>
<td>9.58×10⁻⁴²</td>
<td>2.230 (1.88-2.65)</td>
<td>7.49×10⁻²⁰</td>
<td>2.731</td>
<td>1.59×10⁻⁵⁷</td>
<td>rs10119</td>
</tr>
<tr>
<td>PVRL2</td>
<td>19</td>
<td>rs12972156 (i)</td>
<td>2.942 (2.45-3.54)</td>
<td>0.151</td>
<td>1.52×10⁻³⁰</td>
<td>2.656 (2.06-3.43)</td>
<td>5.42×10⁻¹⁴</td>
<td>2.841</td>
<td>8.30×10⁻⁴³</td>
<td>PVRL2 (rs412776)</td>
</tr>
<tr>
<td>APOC1</td>
<td>19</td>
<td>rs10414043 (g)</td>
<td>3.082 (2.55-3.73)</td>
<td>0.143</td>
<td>3.92×10⁻³¹</td>
<td>2.629 (2.02-3.43)</td>
<td>8.76×10⁻¹³</td>
<td>2.920</td>
<td>4.29×10⁻⁴²</td>
<td>rs10414043</td>
</tr>
<tr>
<td>APOE non-carrier</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CACNA1A</td>
<td>19</td>
<td>rs189753894 (i)</td>
<td>1.726 (1.34-2.23)</td>
<td>0.082</td>
<td>2.68×10⁻⁵</td>
<td>1.900 (1.35-2.67)</td>
<td>2.22×10⁻⁴</td>
<td>1.787</td>
<td>2.49×10⁻⁸</td>
<td>-</td>
</tr>
<tr>
<td>LRIG1</td>
<td>3</td>
<td>rs2280575 (i)</td>
<td>0.539 (0.41-0.71)</td>
<td>0.092</td>
<td>1.15×10⁻⁵</td>
<td>0.551 (0.40-0.76)</td>
<td>3.46×10⁻⁴</td>
<td>0.544</td>
<td>1.51×10⁻⁸</td>
<td>-</td>
</tr>
<tr>
<td>APOE e4 carrier</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* m: missense, s: synonymous, i: intronic, g: intergenic, 3': 3' UTR; †SNP alternative allele frequencies of Europeans reported in the allele frequency aggregator (ALFA) [37]
Figure 1. Gene expression of CACNA1A across different tissues (B) and expression quantitative trait loci (eQTL) analysis of rs2280575 in LRIG1 (A). (A) The CACNA1A is expressed at a high level in the cerebellum (median TPM: 304.8) and in the cerebellar hemisphere (median TPM: 275.0). (B) Violin plot showing the effect of the eQTL rs2280575 on LRIG1 expression in hypothalamus (P = 1.13×10⁻²). (Image available from https://gtexportal.org/home/gene/CACNA1A).