February 2022
Ph.D. Dissertation

Identification of genetic loci associated with late-onset Alzheimer's disease from East Asian genome-wide association study

Graduate School of Chosun University

Department of Integrative Biological Sciences

Sarang Kang

Identification of genetic loci associated with late-onset Alzheimer's disease from East Asian genome-wide association study

전장유전체 연관분석을 통한 동아시아인 특이
알츠하이머병 연관 유전변이 발굴 연구

February 2022

Graduate School of Chosun University

Department of Integrative Biological Sciences

Sarang Kang

Identification of genetic loci associated with late-onset Alzheimer's disease from East Asian genome-wide association study

Advisor: Prof. Kun Ho Lee

This dissertation is submitted to the Graduate School of Chosun University in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Science

October 2021

Graduate School of Chosun University

Department of Integrative Biological Sciences

Sarang Kang

This is to certify that the Ph.D. dissertation of Sarang Kang has successfully met the dissertation requirements of Chosun University.

Chairman	Chosun Univ.	Prof. Seok-Jun Kim	
Member	Chung-Ang Univ.	Prof. Jung-Woong Kim	
Member	Chosun Univ.	Prof. Han Yong Lee	
Member	Chosun Univ.	Prof. Jungsoo Gim	
Member	Chosun Univ.	Prof. Kun Ho Lee	

January 2022

Graduate School of Chosun University

CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Identification of genetic loci associated with late-onset Alzheimer's disease from East Asian genome-wide association study

Sarang Kang
Advisor: Prof. Kun Ho Lee, Ph.D.
Department of Integrative Biological Sciences
Graduate School of Chosun University

Alzheimer's disease (AD) is a highly complex neurodegenerative disorder and is regarded as one of major forms of dementia. The AD is clinically characterized by progressive deficits in memory, behavioral problems and cognitive impairments. High genetic heritability (up to 79%) of Late-onset of Alzheimer's disease (LOAD) have contributed to large scale of genetic studies, and more than 20 genetic loci have been discovered, including the Apolipoprotein E (APOE) gene. However, the majority of genetic loci have been identified through European races, and relatively few studies using East Asians to discover genetic factors. Furthermore, the results of East Asian studies have not clearly shown any risk variants in LOAD, except for the *APOE* gene. The focus was on the identification of Asian specific genetic loci associated with AD. In this study, Genome-wide association studies (GWASs) using 2,291 Korean data collected from Gwangju Alzheimer's & Related Dementias (GARD) Cohort were performed to find the risk genetic loci of AD. The novel genes were identified by performing APOE-strata analysis to effectively mask the genetic effects of *APOE*. And those genes were confirmed the replication using independent Japanese data and found the East Asian specific loci through meta-analysis. The previously reported SNP (or gene) were replicated using GARD cohort and three

novel susceptible loci were discovered through the APOE-stratified analyses. Among subjects without the APOE e4 allele, we identified that rs189753894 in *CACNA1A* gene and rs2280575 and rs113765515 in *LRIG1* gene were susceptible loci for Alzheimer's disease. Next, 115 AD-associated genes were identified by bootstrap-based GWAS on a Korean population followed by functional genomic screening with a Drosophila model of AD. Pathway analysis of the discovered genes revealed that membrane trafficking is the main functional cluster. Among them, the function of 5 selected genes (*AP3D1*, *DNM3*, *SH3GL2*, *PIK3C3*, and *CCDC22*) were validated in vivo and in vitro. Next, the association and risk effect of AD in Koreans for APOE e2 was evaluated. The association analysis confirmed that the risk of AD was 0.39 times lower in the group with at least one APOE e2 than in the group with APOE e3/e3, and that this was more clearly observed in the group with a pathological diagnosis. Moreover, through survival analysis, the group with more than one APOE e2 confirmed a 3-year delay in the age of onset of AD compared to APOE e3/e3. The accumulation of amyloid-beta as age progressed was examined, and it was confirmed that those who had at least one APOE e2 had no or less amyloid accumulation than those who did not. To conclude, several new genetic loci have been identified for the association with AD. Moreover, the East Asian specific risk SNPs, which involved in the AD mediated pathology were also identified. In addition, the findings of this study can provide new insights into understanding the complex of neurodegenerative disease.

I. INTRODUCTION

I-1. Background of Alzheimer's disease dementia

Dementia is a symptom of cognitive decline caused by a variety of diseases, with Alzheimer's disease (AD) being the most common causative disease. With the increase in the elderly population, the number of dementia patients is also increasing. According to the 2016 Korea Dementia epidemiological survey, the estimated number of dementia patients among the elderly aged 65 and above will be approximately 790,000 in 2019, and is expected to exceed 1.36 million in 2030, 2.2 million in 2040, and 3 million in 2050 (Nam et al., 2017). The cost of dementia management is also increasing; in 2010 the cost of management per person with dementia was about 18.5 million won, and is estimated to be 20.7 million won in 2019 (Lee et al., 2021). AD accounts for an estimated 60-80% of all dementia patients, and develops in elderly people over the age of 65. According to the 2020 Korea Dementia Observatory Report, in 2021, among the estimated dementia patients aged 65 years and older, AD patients accounted for the highest estimated at 74.9%, vascular dementia with 8.7% and other types of dementia with 16.3% (Lee et al., 2021).

The most common early symptom of AD is changes in the brain regions responsible for learning and memory, which typically make it difficult to form new memories and remember recently learned information. Serious memory loss, confusion and disorientation, changes in mood and behaviour, and deeper confusion about events, time, and place are clear features of Alzheimer's disease. In AD patients, pathological changes are observed in the brain: senile plaque, neurofibrillary tangle, and brain atrophy. Senile plaques are produced by the accumulation of beta-amyloid protein on the external surface of neurons, while neurofibrillary tangles are the result of hyperphosphorylated tau protein on the internal surface of neurons. Small accumulations of beta-amyloid, called plaques and

oligomers, can contribute to neuronal damage and death by interfering with communication between neurons at the synapse. Tau tangles inside the neuron block the transmission of nutrients and other molecules essential for normal function and neuron survival. Although the entire sequence of events is unclear, beta-amyloid begins to accumulate prior to abnormal tau, and increased accumulation of beta-amyloid is associated with subsequent increases in tau (Bloom, 2014). In AD, the accumulation of beta-amyloid and phosphorylation of tau protein causes the death of neurons throughout the brain, severing the neuronal network and reducing many brain regions. Hippocampus, entorhinal cortex, subcortical structures, and white matter were observed in a large part of the brain, and the volume of the brain was greatly reduced (Pini et al., 2016).

According to the U.S. Dementia Epidemiology Study, the prevalence of AD is different by ethnicity, and even within the same ethnic group, the prevalence is different by environment (Anderson et al., 2004). The CHAP study revealed that among the geriatric population aged 65 and older, 18.6% of Blacks and 14% of Hispanics have Alzheimer's disease, while only 10% of Whites have AD (Rajan et al., 2021, Thies and Bleiler, 2021). This suggests that the factors that cause AD are different for each ethnic group. And more studies, especially cohort-based studies, are needed to derive conclusions about the prevalence and causes of AD.

I-2. Genetic study of AD

Genetic research on genes that cause late-onset AD (LOAD), which occurs at ages over 65, has been conducted since the early 1990s. The high heritability of AD (up to 79%), estimated by twin studies, stimulated genetic research to discover the genes that induce AD (Gatz et al., 2006). And the genome-wide association study (GWAS) accelerated the finding genes associated with AD. The early genome-wide studies led to the identification of many different AD-related genes, such as *CD33, ABCA7, BIN1, CLU, EPHA1* and *SORL1* (Harold et al., 2009, Hollingworth et al.,

2011, Miyashita et al., 2013). GWAS mainly uses single nucleotide polymorphism (SNP), which is the part of the genome where differences appear in each individual's genetic sequence, to investigate associations (Shastry, 2002). SNPs can cause differences in gene expression or changes in protein structure depending on their genomic loci. In addition, the type and number of SNPs vary among races, which can lead to different disease traits. Numerous genetic variants may represent diverse and complex traits, and have synergistic effects on the characteristics of many diseases.

Unfortunately, genes identified through GWAS could not explain even half of the disease mechanisms. Perry G. Ridge et al. reports that through studies of AD heritability estimates using genetic information, only about 33% of heritability was explained, while the remaining 67% was not (Ridge et al., 2013). This is either due to the large number of genes that have not yet been identified or due to biological mechanisms caused by complex combinations of genes.

Recently, researchers have been attempting to discover new genetic variants not only through simple association analysis but also through applied analysis using pathway and protein information (Mukherjee et al., 2014, Zhou et al., 2019, Sun et al., 2021). Not only that, but there are also association studies that increase the sample size to make it more representative of the population. A recent report of association analysis using more than 1 million sample showed the results of adding 7 new genes to the previously reported 32 genes (Wightman et al., 2021). However, most of the reported results have been generated from the research using Europeans, and the results using East Asians are relatively insufficient.

I-3. Genetic study on apolipoprotein E (APOE)

The most genetically established LOAD risk factor is the e4 allele of the Apolipoprotein E (APOE) gene on chromosome 19. *APOE* is a gene that codes for proteins involved in fat metabolism, such as cholesterol transport in neurons (Martins

et al., 2006). The human *APOE* gene has three isotypes e2, e3, and e4 depending on the combination of two SNPs (rs429358 and rs7412), with frequencies of 8.4%, 77.9%, and 13.7%, respectively (Farrer et al., 1997). Since human genome is diploid, *APOE* allele can be inherited from each parent, resulting in six allele types: e2/e2, e2/e3, e2/e4, e3/e3, e3/e4 and e4/e4. The researchers observed a difference in the frequency of such pairs by race and ethnicity. For example, comparing populations with APOE e4/e4, African Americans were 4.5% compared to 2.4% of European Americans. For the population with APOE e3/e4, the percentage of African Americans was 28.6% compared to 21.4% for European Americans (Rajan et al., 2017).

The frequency of the e4 allele is greatly increased in AD patients. Having one copy of the e4 allele of APOE increases the risk of AD 4-fold, and having two copies increases it more than 10-fold (Farrer et al., 1997). In addition, amyloid-beta accumulation occurs more in people who have APOE e4 than in those who do not (Jansen et al., 2015). And people who have APOE e4 accelerate the onset of AD compared to those who have APOE e3 (Raber et al., 2004). A number of GWAS studies have confirmed that APOEe4 is the most powerful genetic risk factor for AD (Shen and Jia, 2016).

The APOE e2 has the lowest frequency in populations worldwide, especially in patients with AD. The e2 allele of APOE provides a protective effect against AD (Corder et al., 1994). When compared to the APOE e3/e3 population, which had the highest frequency, the population with APOE e2 had the lowest risk of AD and the highest mean age (Corder et al., 1994, Farrer et al., 1997). The protective effect of ε2 does not vary by ethnicity, age or sex (Farrer et al., 1997).

PART II

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

SUMMARY

The high genetic heritability of Alzheimer's disease (AD) (up to 79%) has stimulated genetic studies for AD. Genome-wide association studies (GWAS) using consolidative procedures, large sample sizes and meta-analysis have identified more than 30 high-risk genetic loci for AD. However, most studies have been conducted on Caucasians, and studies using East Asians have been relatively insufficient. Therefore, it is not enough to understand the genetic mechanisms of AD, which are genetically complex, heterogeneous, and ethnically distinct. In this study, GWAS using 2,291 Korean data collected from Gwangju Alzheimer's & Related Dementias (GARD) Cohort was performed to discover the genetic risk factors for AD. As shown in other studies, 54 SNPs at or near *APOE* displayed genome-wide significance, while 432 SNPs remained at a suggestive level. To completely mask the genetic effects of *APOE*, novel genetic loci were identified through *APOE* strata analysis. Then, the replication study was performed using independent Japanese data, and East Asian-specific loci were confirmed through meta-analysis. The previously reported SNPs (or genes) were replicated using the GARD cohort, and three novel susceptibility loci were identified through APOE-stratified analysis. Among subjects without the APOE e4 allele, rs189753894 in *CACNA1A* gene and rs2280575 and rs113765515 in *LRIG1* gene were identified as susceptible loci for AD. This study shows that detection of AD-associated variants is possible in ethnic groups with more homogeneous genetic backgrounds, using a sample composed of fewer subjects.

II-1. Introduction

Genetic factors are consistently regarded as important predictors in understanding the mechanisms of Alzheimer's disease (AD). Among the genetic factors associated with AD, the most critical and consistently reported gene is Apolipoprotein E (APOE), which explains the largest portion of the disease (Strittmatter et al., 1993). However, differences in the effects of APOE genotype are highly variable among populations (Singh et al., 2006). A notable example is that the prevalence of AD among Arabs living in communities in northern Israel is almost double that of their European ancestors, even though the frequency of APOE e4 is very low (Sherva et al., 2011). In 2013, 11 new AD genetic variants were discovered through a meta-analysis using 74,046 individuals (Lambert et al., 2013). The International Genomics of Alzheimer's Project (IGAP) in 2019, 24 AD-related genes were discovered using 94,437 individuals (Kunkle et al., 2019). Most recently, using more than a 1 million people from all cohorts combined, 39 genes were confirmed to have associations with AD (Wightman et al., 2021). The discovered genes are involved in diverse biological pathways, such as immune responses and endocytosis, and give a potential explanation for the complexity and genetic heterogeneity of AD. However, the large-scale studies that identified the genes used mostly European populations, which is quite different from the results of studies using East Asians (Miyashita et al., 2013, Zhou et al., 2018). This supports the research that the genetic distance between Europeans and East Asians is large (Abdulla et al., 2009, Auton et al., 2009)and further emphasizes the need for research using East Asians.

Genetic studies using diverse populations have been very valuable in understanding the genetic architecture of AD (Miyashita et al., 2013, Ghani et al., 2015, Jun et al., 2017, Jansen et al., 2019). Studies using non-European populations provide an opportunity to reveal new AD-related genes, which may serve as a basis for understanding unknown genetic mechanisms of disease. In particular, studies

using a variety of populations have found genes that explain genetic heterogeneity, even though the populations consisted of thousands or far fewer samples (Meng et al., 2006, Jun et al., 2017).

In this study, the genetic information of Koreans, a homogeneous population that maintains a well-defined genetic background with a high prevalence of AD among the elderly over 60 years old, was used to discover AD-related genes. For this purpose, the genetic information for conducting the GWAS was obtained using the Korea Biobank Array (referred to as Korean Chip) designed for Koreans (Moon et al., 2019). In order to extend the genetic information, imputation using a multi-ethnic panel was conducted (McCarthy et al., 2016). The SNPs discovered through the Koreans were validated through Japanese, the most genetically homogeneous population (Miyashita et al., 2013).

II-2. Methods

II-2-1. GARD cohort

GARD (Gwangju Alzheimer's & Related Dementias) cohort is a group that has been tracked since 2010 after recruiting local senior citizens aged 60 or older for research on Alzheimer's disease and dementia. Most of the subjects are recruited at the Early Dementia Examination Center in Bitgoeul Senior Health Town, Gwangju. Basic personal information such as height, weight, education, medical history, family history, and drug use are collected. Biometric tests (EEG, PPG, impedance) and K-MMSE (Korean Mini-Mental State Examination) tests (Kang et al., 1997) are performed to assess basic cognitive functions and to determine whether dementia is judged or not. Currently, the number of subjects who have been assessed for MMSE is 14,524. After conducting the basic tests, additional full clinical examinations are conducted for those who wish to have a detailed examination or who agree to the investigation. A clinical evaluation was conducted using the Seoul

Neuropsychological Screening Battery (SNSB) (Kang et al., 2003) and brain imaging information using a 3T MRI (Skyra, Siemens) single scanner. ^{18}F-Florbetaben (FBB) beta-amyloid positron emission tomography (PET) was conducted on some subjects who wanted and needed. Based on all the clinically collected information, dementia specialists in the Department of Neuropsychiatry at Chosun University Hospital and Chonnam University Hospital conducted the diagnosis. The Blood, saliva, and cerebrospinal fluid (CSF) were collected from the diagnosed subjects. After the blood and CSF are collected, they are immediately purified and stored in deep freezer and nitrogen tank. A portion of the blood collected was tested for blood counts, cholesterol levels, blood sugar, thyroid function, and other blood information. The cognitively normal (CN) subjects had no neurological diseases or disorder, no observed decline in cognitive function, and no problems in daily life. Subjects with MMSE z-scores greater than -1 or with less than 3 years of education were excluded from the CN group. Subjects with brain disease, central nervous system disease, or mental health disorders, including depression, were excluded from the CN group. Those taking stroke medications, mental health related medications including sleeping pills, and those drinking alcohol more than four times a week were also excluded from the group. The clinical diagnosis of probable AD was determined using NINCDS/ADRDA criteria (McKhann et al., 1984). Amyloid PET images were used to diagnose AD. All volunteers or authorized guardians for cognitively impaired individuals gave written informed consent before participation.

II-2-2. Study participants

Out of 14,524 participants, the subjects were chosen to maximize contrast between cases and controls for the GWAS. The cases had an age distribution of 60 to 97 years and met the NINCDS-ADRDA for AD. Those with amyloid negative diagnosis by PET imaging were excluded. The control group consisted of individuals whose age distribution ranged from 70 to 90 years and who were cognitively normal

or did not exhibit neurological, psychological, or pathological symptoms. For the expansion of hospital-based samples, blood or DNA with clinical information was received from 13 institutions, including Seoul National University Hospital, Inha University Hospital, Gyeongbuk University Hospital, Dong-A University Hospital, and Pusan National University Hospital. The study protocol was approved by the Institutional Review Board of Chosun University Hospital, Korea (CHOSUN 2013-12-018-070).

Japanese datasets were used to performed a replication study to identify the novel East Asian specific genes. Professor Takechi Ikeuchi of Niigata University in Japan gave clinical information and genetic data for a total of 2,022 Japanese cases and controls (Miyashita et al., 2013).

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

	Discovery (n=2,291)		P	Replication (n=1,956)		P
	CN	AD		CN	AD	
All samples, n	1,172	1,119		976	980	
Female, n (%)	661 (56.4)	715 (63.9)	<0.001	564 (57.8)	702 (71.6)	<0.001
Age at exam, m (s.d)	76.03 (8.9)	74.60 (8.9)	<0.001	76.94 (5.9)	72.99 (4.3)	<0.001
APOE e4 non-carrier, n	976	621		815	435	
Female, n (%)	544 (55.7)	393 (63.3)	0.003	470 (57.7)	31 (71.3)	<0.001
Age, m (s.d)	76.10 (8.9)	75.43 (9.1)	0.029	77.11 (5.9)	73.42 (4.3)	<0.001
APOE e4 carrier	196	498		161	545	
Female, n (%)	117 (59.7)	322 (64.7)	0.222	94 (58.4)	392 (68.3)	0.001
Age, m (s.d)	95.45 (8.9)	73.32 (9.0)	<0.001	76.06 (5.9)	72.65 (4.3)	<0.001

Abbreviations: CN, cognitive normal; AD, Alzheimer's disease; s.d, standard deviation.

II-2-3. SNP genotyping

Genomic DNA from GARD cohort was extracted from peripheral blood leukocytes that were isolated from whole blood collected in EDTA tube. Of the total subjects, 5,570 blood samples were finally used for genomic analysis, except for diagnostic uncertainty, mixed-up, and inadequate DNA concentrations among all subjects. The samples were genotyped using the KNIH Biobank Array as part of the discovery sample. Affymetrix Power Tools (APT) were used to process all CEL files. Dish QC (DQC) values were generated and used to remove samples with DQC<0.82. Low-quality markers selected by the Ministry of Health to improve marker quality control have been removed. Also, low-quality markers were removed by SNPolisher to pass through a better sample QC procedure. The 4,391 samples used KNIH Biobank array v1.0, while 1,179 samples used KNIH Biobank array v1.1 which is a more supplemented by Koreans. The number of samples analyzed for discovery study was expanded by integrating both versions and extracting common SNPs. The replication data was genotyped with Affymetrix GeneChip 6.0 microarrays, and includes 905,422 variants for 2,022 subjects (1,007 AD cases and 1,015 controls).

II-2-4. Quality control

Quality control was performed under the same conditions for each chip version using the PLINK v1.90 package (Chang et al., 2015). SNPs with genotype call rate (GCR) < 95%, a minor allele frequency (MAF) <0.01, or significant deviation from the Hardy-Weinberg equilibrium (HWE) ($P<10^{-6}$) were excluded. Samples with individual call rate < 95%, gender inconsistency between reported and analysis of X-chromosome SNPs, extremely low or high genome-wide heterozygosity (±3 S.D from the mean). After performing QC for SNPs and samples, 4,011 samples and 518,326 SNPs were passed from Korean chip v1.0 and 1,117 samples and 577,694 SNPs were passed from Korean chip v1.1. For discovery study,

totally 2,291 subjects (1,172 controls and 1,119 AD cases) were selected (Table 1). For replication study, sample QC and marker QC were conducted on the same basis as discovery dataset, and finally 1,956 samples and 669,196 variants were used (Table 1).

II-2-5. Imputation

SNP genotypes for Korean and Japanese were imputed separately using reference haplotypes pre-phased in the HRC (Haplotype Reference Consortium) panel version 1.1 (McCarthy et al., 2016). Eagle (version 2.3) was used for phasing, and Minimac4 was used for imputation. The low-quality (info score <0.5) imputed SNPs were removed. After imputation, 35,685,761 variants were obtained by combining the data from Korean discovery dataset and 39,044,005 variants from Japanese replication dataset.

II-2-6. Statistical analyses

Genetic Association Study (GAS) Power Calculator was used to analyze the power of GAWS (Johnson and Abecasis, 2017). Demographic information with the statistical significance was confirmed through Chi-square test or t-test (Table 1). The allelic test or a logistic regression procedure was performed using PLINK v1.90 to test both genotyped and imputed SNPs for association (Chang et al., 2015). Under the additive gene model, the logistic regression analyses were adjusted for age, sex and the first four principal components (PCs), suggested from the smartpca program with EIGENSTRAT (Price et al., 2006, Patterson et al., 2006). Using twstat software in EIGENSTRAT, Tracy-Widom statistics were calculated to evaluate the statistical significance of each PCs, and the first 4 PCs for discovery stage and 5 PCs for replication stage were included in this study.

To mask the high effect of APOE signal, APOE strata analysis was used. To perform the APOE-stratified GWAS, each group was clustered according to the

presence or absence of APOE e4. A logistic regression model stratified on the presence of APOE genotypes, especially e4, was configured to search for effect sizes. SNPs with suggestive significance ($P<5\times10^{-5}$) in discovery stage were selected as the candidates and considered significance in replication stage.

In replication stage, 1,956 samples were analyzed similarly done in the discovery and replication stage, except 5 PCs were adjusted. The association of p value less than 1×10^{-3} was considered significance. The SNPs with reverse direction of the odds ratio were excluded.

The meta-analysis was performed using METAL (Willer et al., 2010), and SNPs were selected with p less than 5×10^{-5}. An inverse variance-weighted fixed effects meta-analysis was used to test for association with AD. And gene-set analysis was also conducted through MAGMA for top SNPs (de Leeuw et al., 2015). GCTA software was used to estimate the SNP heritability as defined by phenotypic variance explained by genotype information (Yang et al., 2011).

II-2-7. Gene expression analysis

The expression levels of mRNA and protein in normal human tissues were considered using public databases that provide open resources, such as Genotype-Tissue Expression (GTEx, release v8) (https://gtexportal.org/home/) (Consortium et al., 2015) and Human Protein Atlas (HPA) (http://www.proteinatlas.org/) (Uhlén et al., 2005), and Functional Annotation of Mammalian Genomes 5 (FANTOM5) (https://fantom.gsc.riken.jp/5/) (Lizio et al., 2015).

Genotype-specific expression of SNPs was assessed by quantitative characteristic position of expression (eQTL) analysis using GTEx. The marker ID for rs189753894 in *CACNA1A* was chr19_13513675_C_A_b38. The marker ID for rs2280575 in *LRIG1* was chr3_66492439_A_G_b38. The marker ID for rs113765515 in *LRIG1* was chr3_66489422_T_G_b38.

In addition, the publicly available expression dataset (GEO access number: GSE118553), which specifically contains AD data, was used to perform the differential expression (DE) analysis (Patel et al., 2019). The DE dataset was included expression levels of gene in brain tissue from 167 AD patients (both asymptomatic and AD dementia) and 100 cognitive normal controls. The linear regression model adjusting for age and sex was performed to examine the differences among disease status.

II-3. Results

II-3-1. Allelic association results using all samples in discovery stage

As a result of allelic association test (chi-square statics), 70 loci (6 genes) were satisfied genome-wide significance level ($P<5\times10^{-8}$) (Table 2). All loci with genome-wide significance were in the previously reported AD risk genes including *APOE, PVRL2, TOMM40, and APOC1*. Among the 479 loci (98 genes) with suggestive levels, SNPs in *BIN1* and *ABCA7* genes, which were previously reported as AD risk genes, were found (Table 2). This suggests that there is possibility of identifying AD risk loci even at the suggestive level.

II-3-2. GWAS for Alzheimer's disease using Korean cohort

For discovery stage, association analysis was conducted using the all samples, which included 2,291 subjects and identified 54 genome-wide significant ($P<5\times10^{-8}$) SNPs across three genes namely *APOE*, *PVRL2*, and *TOMM40* (genomic inflation=1.03), which were found to be associated with AD in previous studies (Table 3 and Figure 1). The heritability of AD was estimated without these genes using a liability-threshold model with the set of SNPs excluding the loci annotated in these genes; and a relatively large estimate (0.566±0.152) was obtained (Table 4). In addition, a portion of SNPs in other previously known genes, such as *ABCA7* and

BIN1, was also replicated at a genome-wide suggestive level ($P<5\times10^{-5}$) in association test using additive model (Figure 1). These results suggest that unreported additional common variants with smaller effects (or possibly rarer variants with larger effects) may be identified through a post-hoc analysis with a less stringent statistical significance. In addition, the rs11218343 in *SORL1*, previously identified in a Japanese cohort (Miyashita et al., 2013), was also confirmed in Korean cohort by replication analysis with a significance of $P=5.9\times10^{-4}$.

II-3-3. Identification of novel genes associated with AD using APOE-stratified GWAS

The APOE-stratified GWAS was performed using subgroups of subjects on the basis of APOE e4 carrier status. The SNPs reached at genome-wide suggestive level ($P<5\times10^{-5}$) were selected for candidates. A total of 219 SNPs (annotated in 61 genes) were identified as e4 carriers (genomic inflation=1.02) with statistical significance respectively. Also, total of 306 SNPs (annotated in 82 genes) were identified as e4 non-carriers (genomic inflation=1.03). Lowering the threshold of statistical significance would obviously increase the number of false findings. Therefore, an independent cohort dataset was used to ensure replicability of the discovered SNPs.

To ensure significance for novel discovered SNPs, a replication study was performed with the Japanese dataset in which previous GWAS signals were filtered. The results of the replication study showed that 15 out of 82 genes of APOE e4 carriers satisfied the significance. Not only that, 7 out of 61 genes of APOE non-carriers satisfied the significance. Considering the effect direction and statistical significance of the variants in both the discovery and replication datasets, the rs189753894 in *CACNA1A* and the rs2280575 and rs113765515 in *LRIG1* achieved significance in the meta-analysis using APOE e4 non-carrier (Table 3).

Furthermore, the power analysis was conducted to evaluate the expected power of the e4 non-carrier findings in discovery stage. Considering SNPs in LD (D'=1) with a prevalence of dementia of 0.087 and a risk allele frequency of 0.082, the power was estimated at 85% with a significant association of $P<5\times10^{-5}$ from the additive model with powerful effect of OR=1.8. In addition, the Allele Frequency Aggregator (ALFA) (www.ncbi.nlm.nih.gov/snp/docs/gsr/alfa/) (Phan et al., 2020) was applied to check the allele frequency of Europeans. Considering that the frequency difference of some loci more than doubles between Koreans and Europeans (Table 3), it suggests that ethnic differences appear even with the same risk loci of AD.

II-3-4. Functional annotation of *CACNA1A* and *LRIG1*

The consortium databases such as HPA (Uhlén et al., 2005), GTEx (Consortium et al., 2015), and FANTOM5 (Lizio et al., 2015) was used as data to support their molecular states of cells and tissues. The results of the eQTL analysis using GTEx showed that two SNPs located in *LRIG1* were produced significant expression differences in Hypothalamus of the brain (Figure 6 and Figure 7). The result of the eQTL analysis with rs189753894 in *CACNA1A* was not showed difference due to its low frequency.

The RNA expression levels of *CACNA1A* were showed as only in brain and high score, such as 18.2 protein-coding transcripts per million (pTPM) in cerebral cortex reported by HPA dataset, 394.1 pTPM in cerebellum reported by GTEx, 453.4 scaled tags per million in cerebellum by reported FANTOM5 dataset (Figure 9). The RNA expression levels of *LRIG1* are showed as high score in the cerebral cortex at 74.1 protein-coding transcripts per million (pTPM) reported by HPA, 16.0 pTPM reported by GTEx. In FANTOM5 dataset, the RNA expression levels of LRIG1 were higher in the brain region than other tissues, especially for the hippocampal formation, which was 179.9 scaled tags per million.

From the DE analysis, the *LRIG1* gene caused significant expression differences in Entorhinal cortex, Frontal cortex, Temporal cortex (Figure 10). Especially when comparing with the CN group, the expression levels clearly changed as the dementia progressed. For the *CACNA1A* gene, a significant difference in gene expression was observed only in the Entorhinal cortex (P=0.0017) (Figure 10).

Table 2. Top-ranked genome-wide allelic association results in discovery stage ($P<5\times10^{-5}$)

Chr	Lead SNP (type)*	Allele	Count	Genotype frequency		Gene	P	OR (95% CI)
				Case	Control			
All samples								
19	rs429358 (m)	CC	78	0.257	0.084	APOE	5.84×10^{-55}	3.76 (3.2-4.5)
		CT	420					
		TT	621					
19	rs10119 (3')	AA	84	0.269	0.096	TOMM40	1.31×10^{-52}	3.48 (2.9-4.1)
		AG	434					
		GG	601					
19	rs10414043 (g)	AA	55	0.212	0.078	APOC1	4.07×10^{-49}	3.33 (2.8-3.9)
		AG	364					
		GG	700					
19	rs111789331 (g)	AA	77	0.256	0.096	APOC1P1	2.96×10^{-49}	3.24 (2.7-3.8)
		AT	644					
		TT	1570					
19	rs412776 (i)	AA	77	0.252	0.108	PVRL2	1×10^{-36}	2.77 (2.4-3.3)
		AG	409					

		GG	633					
2	rs35103166 (g)	CC	538	0.507	0.445	BIN1-NIFKP9	3.27×10^{-5}	1.28 (1.1-1.4)
		CT	1102					
		TT	651					
19	rs3752243 (s)	GG	319	0.354	0.413	ABCA7	3.48×10^{-5}	0.78 (0.7-0.9)
		GA	1123					
		AA	849					
APOE e4 non-carrier								
19	rs189753894 (g)	AA	8	0.109	0.066	CACNA1A	1.55×10^{-5}	1.74 (1.4-2.2)
		AC	119					
		CC	494					
3	rs113765515 (i)	GG	1	0.064	0.105	LRIG1	2.02×10^{-5}	0.56 (0.4-0.7)
		GT	74					
		TT	546					
APOE e4 carrier								
-	-	-	-	-	-	-	-	-

Abbreviations: Chr, chromosome; SNP, single nucleotide polymorphism; OR, odds ratio; CI, confidence interval.

*m: missense, s: synonymous, i: intronic, g: intergenic, 3': 3' UTR.

Table 3. Top-ranked genome-wide association results in discovery stage ($P<5\times10^{-5}$), replication stage, and meta-analyses.

Chr	SNP (type)*	m	MAF	Gene	Discovery (1,172/1,119)		Replication (976/980)		Meta-analysis		Alt Allele Frq†
					OR (95%CI)	P	OR (95%CI)	P	OR	P	
	All samples										
19	rs429358 (m)	C	0.169	APOE	3.637 (3.03-4.37)	3.74×10^{-43}	2.608 (2.04-3.33)	1.89×10^{-14}	3.227	5.90×10^{-55}	0.036
19	rs10119 (3')	A	0.180	TOMM40	3.378 (2.83-4.03)	9.58×10^{-42}	2.230 (1.88-2.65)	7.49×10^{-20}	2.731	1.59×10^{-57}	0.279
19	rs12972156 (i)	G	0.151	PVRL2	2.942 (2.45-3.54)	1.52×10^{-30}	2.656 (2.06-3.43)	5.42×10^{-14}	2.841	8.30×10^{-43}	0.001
19	rs10414043 (g)	A	0.143	APOC1	3.082 (2.55-3.73)	3.92×10^{-31}	2.629 (2.02-3.43)	8.76×10^{-13}	2.920	4.29×10^{-42}	0.109

APOE non-carrier											
19	rs189753894 (i)	A	0.082	CACNA1A	1.726 (1.34-2.23)	2.68×10^{-5}	1.900 (1.35-2.67)	2.22×10^{-4}	1.787	**2.49×10^{-8}**	0.007
3	rs2280575 (i)	G	0.092	LRIG1	0.539 (0.41-0.71)	1.15×10^{-5}	0.551 (0.40-0.76)	3.46×10^{-4}	0.544	**1.51×10^{-8}**	0.267
3	rs113765515 (i)	G	0.088	LRIG1	0.541 (0.41-0.71)	1.78×10^{-5}	0.499 (0.35-0.71)	1.03×10^{-4}	0.525	**7.69×10^{-9}**	0.206
APOE e4 carrier											
-	-	-	-	-	-	-	-	-	-	-	

*m: missense, s: synonymous, i: intronic, g: intergenic, 3': 3' UTR; †SNP alternative allele frequencies of Europeans reported in the allele frequency aggregator (ALFA).

Abbreviations: Chr, chromosome; SNP, single nucleotide polymorphism; BP, base-pair position; m, minor allele; M, major allele; MAF, minor allele frequency; OR, odds ratio; CI, confidence interval.

Table 4. SNP heritability estimates of Alzheimer's disease dementia

	Heritability, V(G)/Vp (SE)
Total SNP	0.596 (0.152)
Exclude APOE SNPs (rs429358 and rs7412)	0.596 (0.152)
Exclude APOE and LD SNPs ($r^2 > 0.2$)	0.574 (0.152)
Exclude APOE and near SNPs (±50kb)	0.566 (0.152)

Abbreviations: SNP, single nucleotide polymorphism; SE, standard error

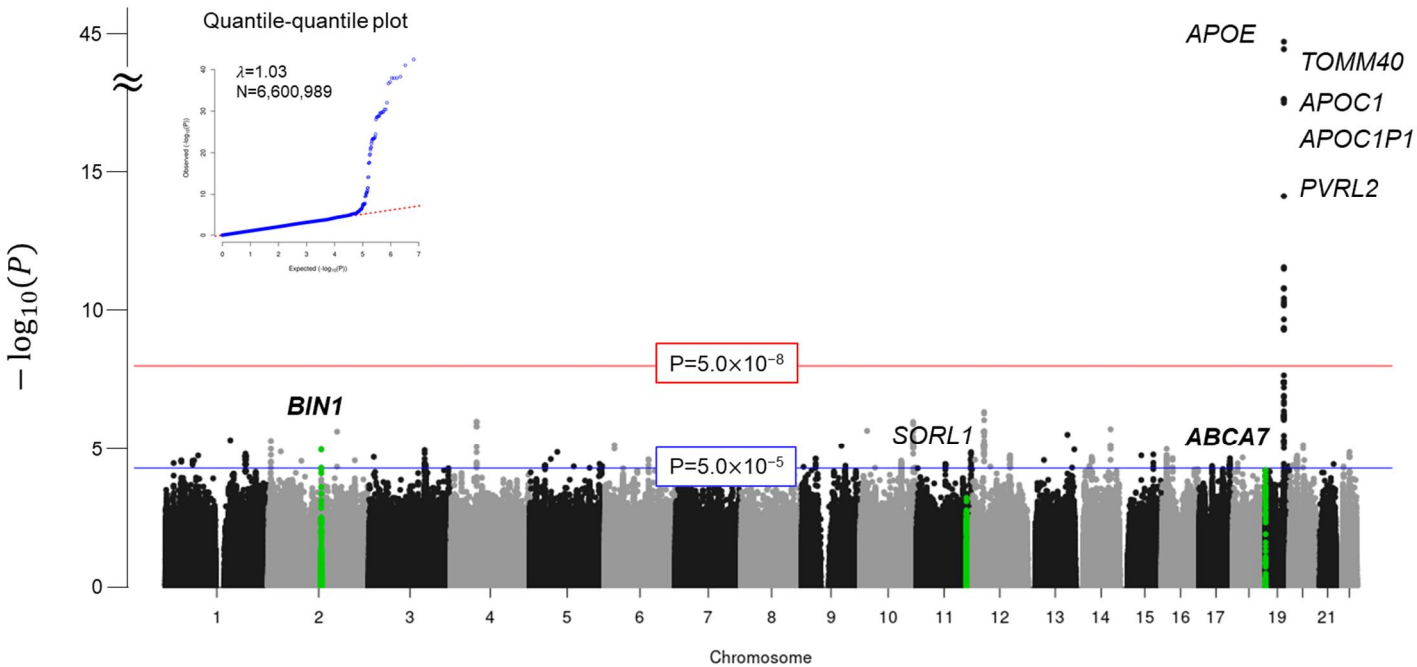


Figure 1. **Manhattan plot and Q-Q plot of GWAS using all samples.** X-axis shows chromosomal positions and Y-axis shows -log10 p values. The upper and lower dotted lines indicate the genome-wide significance threshold ($p=5.0\times10^{-8}$) and the suggestive level for selecting SNPs for replication study ($p=5.0\times10^{-5}$).

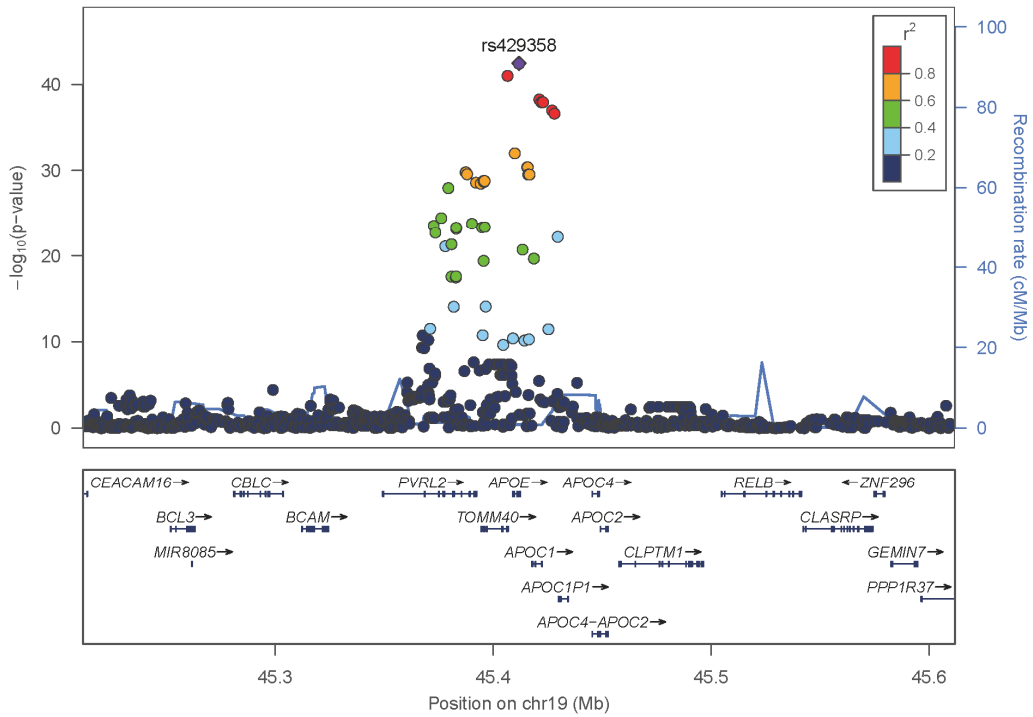
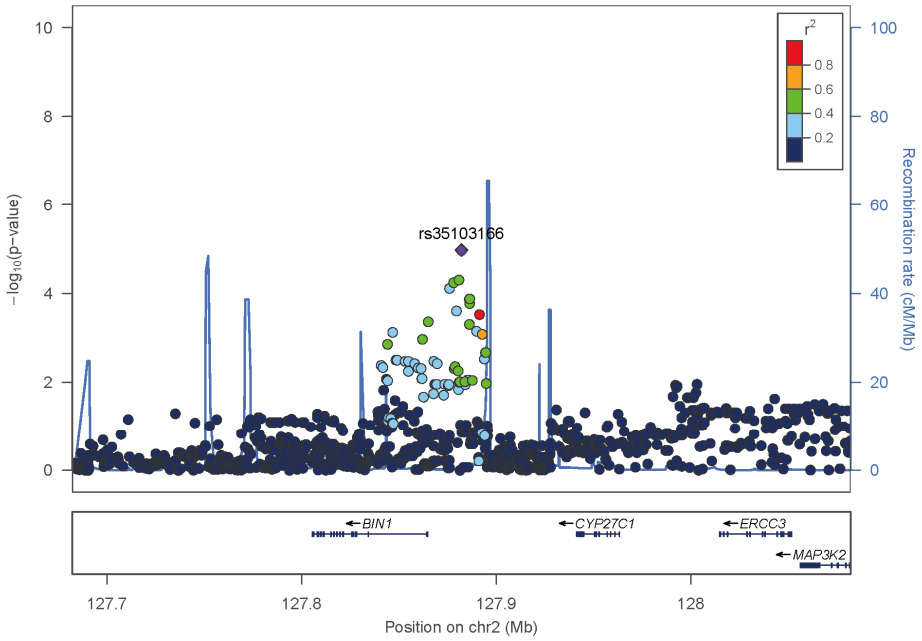


Figure 2. Regional association plots for *APOE* region using all samples. A number of SNPs around rs429358 in *APOE* gene are highly correlated. Region within 200 kb from SNPs showing lowest p value is displayed. SNPs showing the lowest p value is depicted as a purple diamond. Other SNPs colored according to the extent of linkage disequilibrium (measured in r^2) with SNP showing the lowest p value.

A **BIN1**



B **ABCA7**

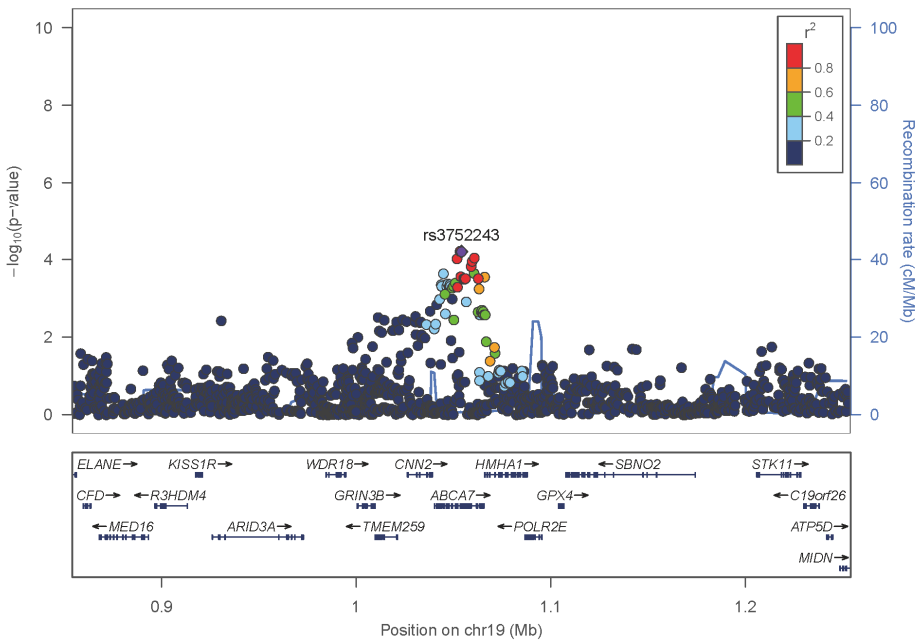


Figure 3. Regional association plots for SNPs that have been previously reported to be associated with AD. BIN1 (A) and ABCA7 (B) were showed the significance with genome-wide suggestive level in Koreans. Region within 200 kb from SNPs showing lowest p value is displayed. SNPs showing the lowest p value is depicted as a purple diamond. Other SNPs colored according to the extent of linkage disequilibrium (measured in r^2) with SNP showing the lowest p value.

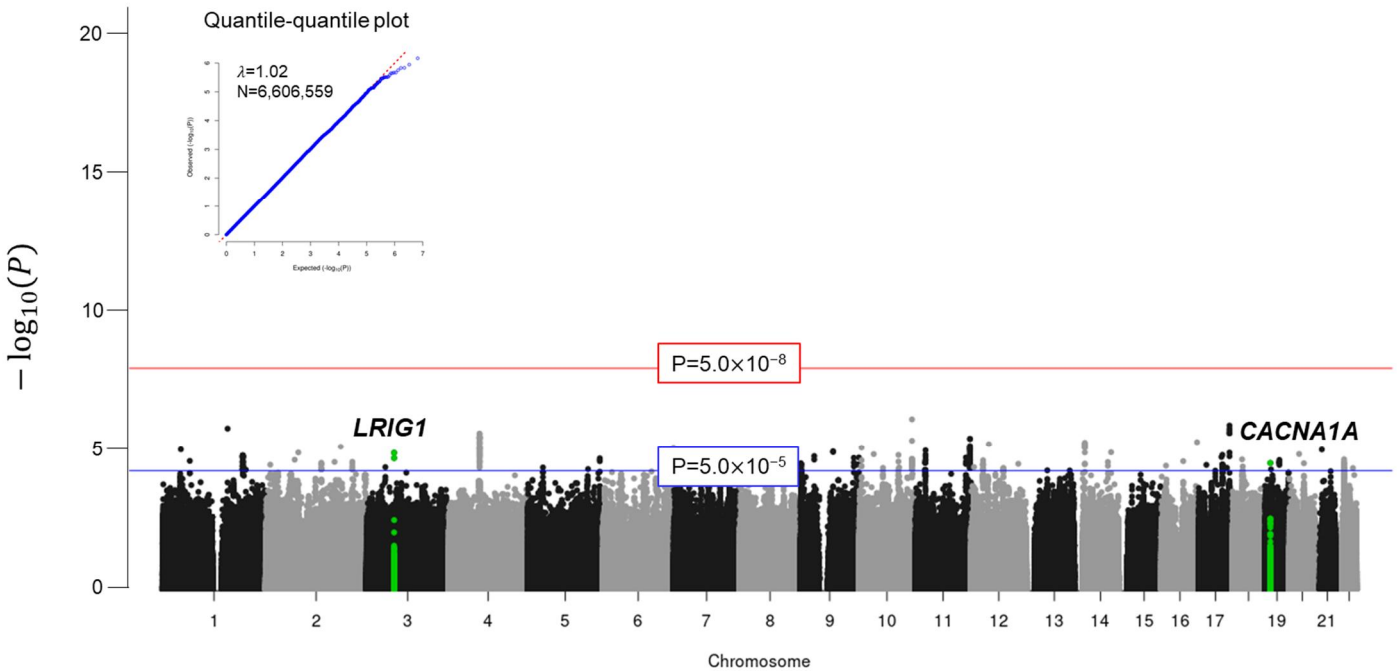


Figure 4. **Manhattan plot and Q-Q plot of GWAS using APOE e4 non-carriers.** X-axis shows chromosomal positions and Y-axis shows -log10 p values. The upper and lower dotted lines indicate the genome-wide significance threshold ($p=5.0\times10^{-8}$) and the cut-off level for selecting single-nucleotide polymorphisms for replication study ($p=5.0\times10^{-5}$).

A **LRIG1**

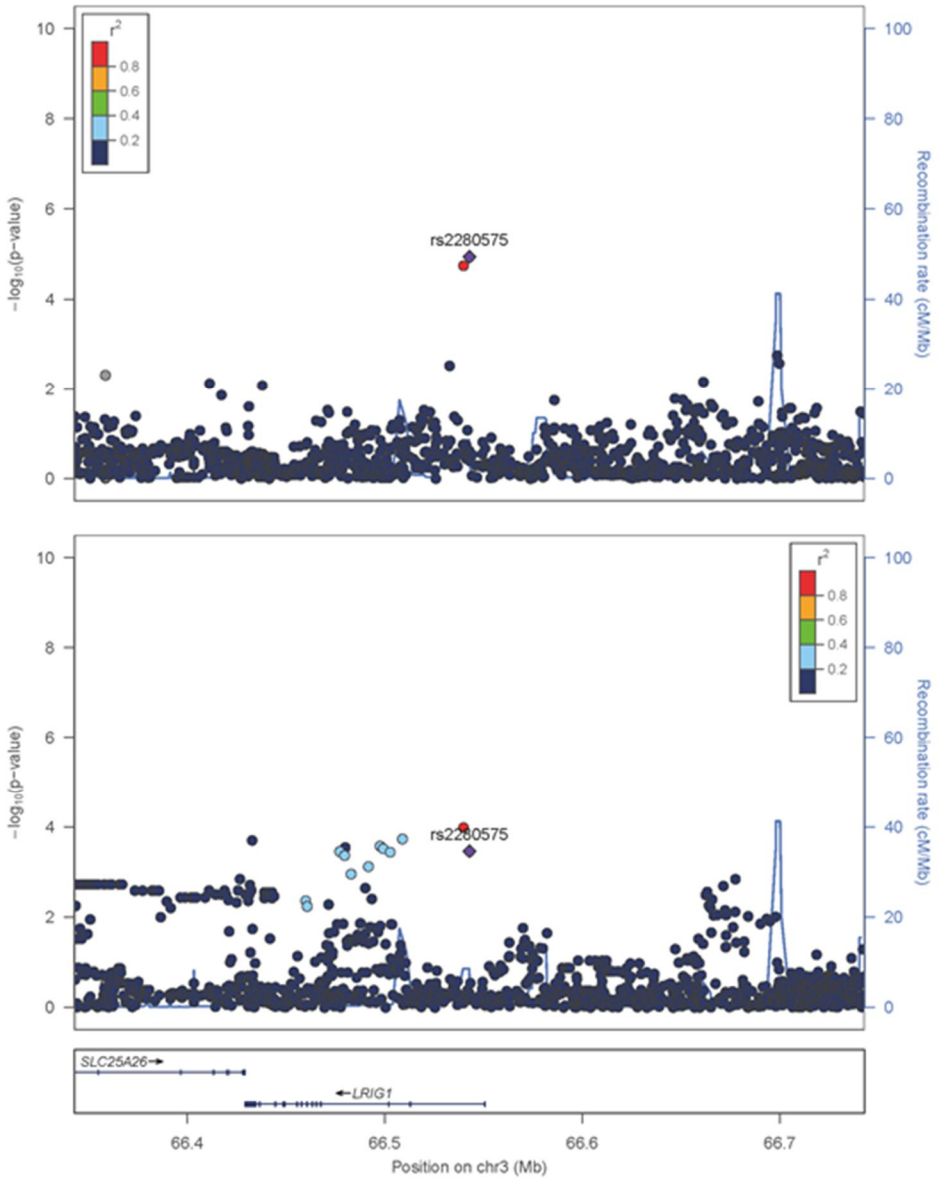


Figure 5. Continued

B **CACNA1A**

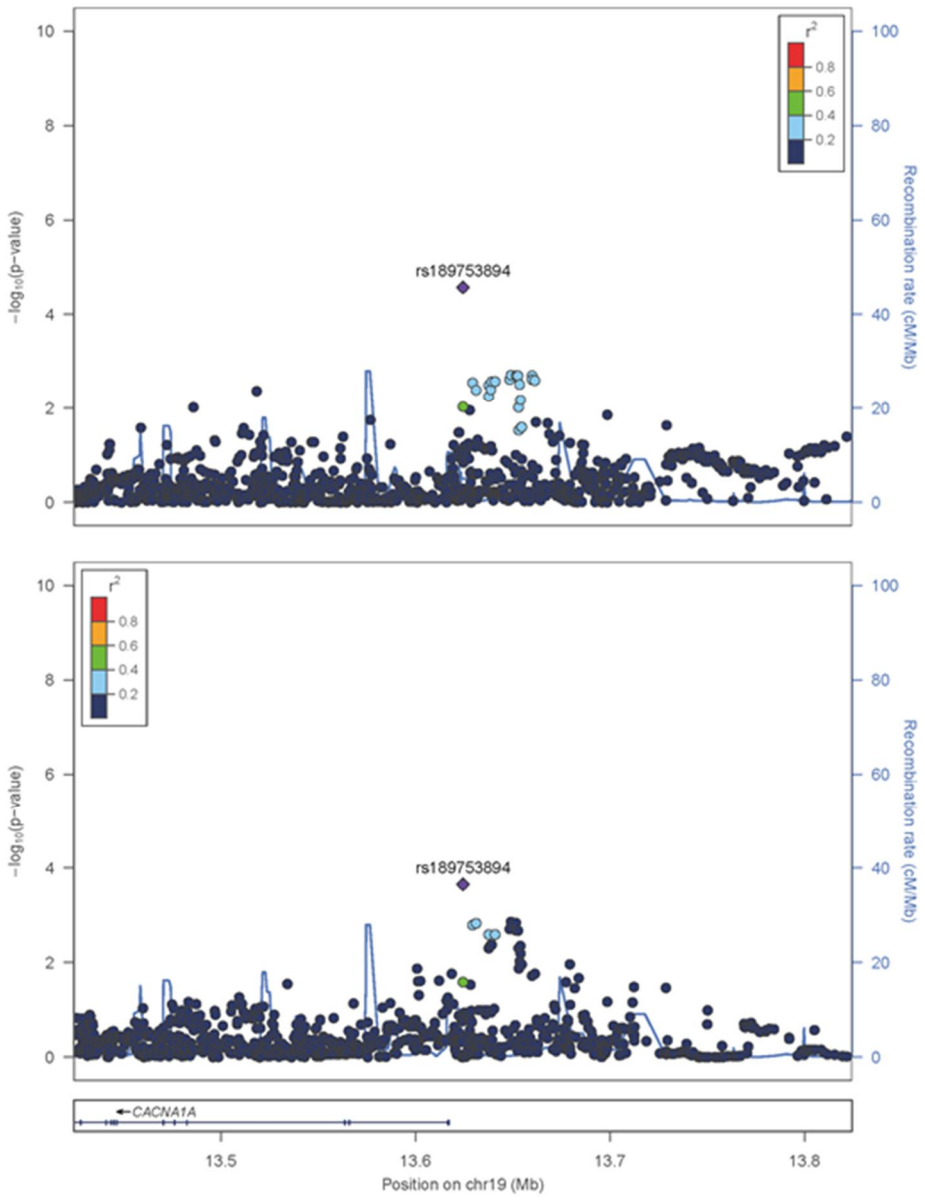


Figure 5. Regional association plots for top-ranked SNPs from APOE e4 non-carrier study. Each plot of SNPs has a Korean panel at the top and a Japanese panel at the bottom. The highest association signal in each panel is located on *LRIG* (A), *CACNA1A* (B). Region within 200 kb from SNPs showing lowest p value is

displayed. SNPs showing the lowest p value is depicted as a purple diamond. Other SNPs colored according to the extent of linkage disequilibrium (measured in r^2) with SNP showing the lowest p value.

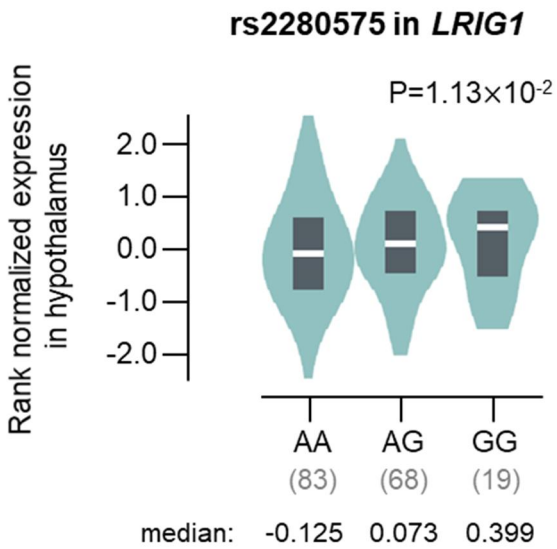


Figure 6. Violin plot showing the effect of the eQTL rs2280575 on *LRIG1* expression in the hypothalamus ($p=1.13\times10^{-2}$).

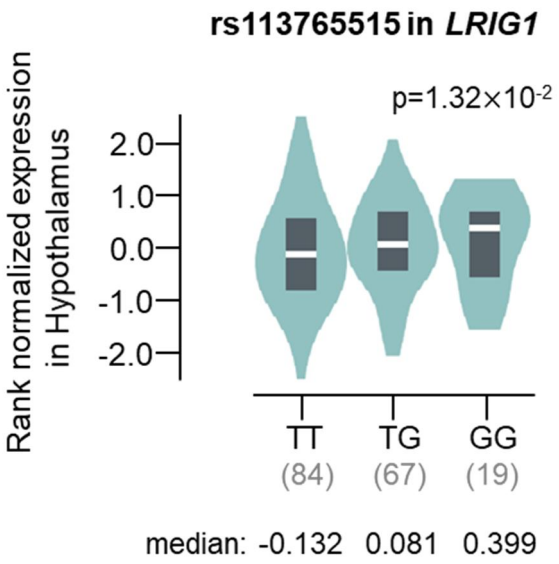


Figure 7. Violin plot showing the effect of the eQTL rs113765515 on *LRIG1* expression in the hypothalamus ($P=1.32\times10^{-2}$).

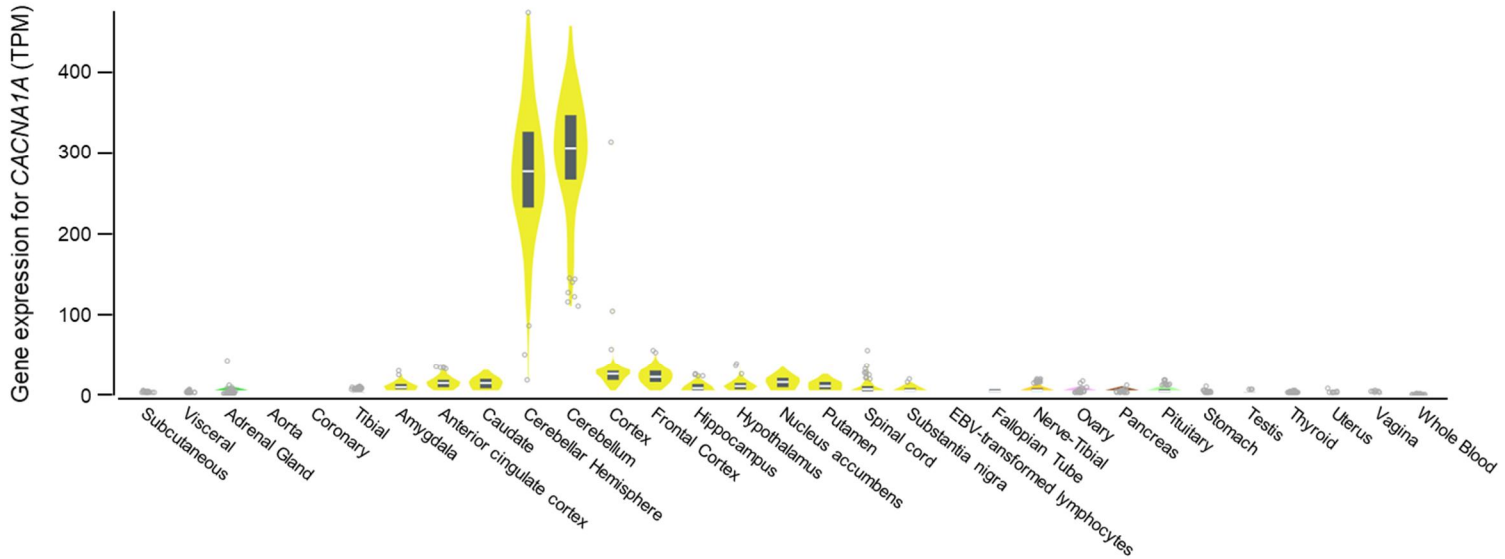


Figure 8. The expression level of two genes that show significant differences by tissue. (A) The CACNA1A is expressed at a high level in the cerebellum (median TPM: 304.8) and in the cerebellar hemisphere (median TPM: 275.0).

A

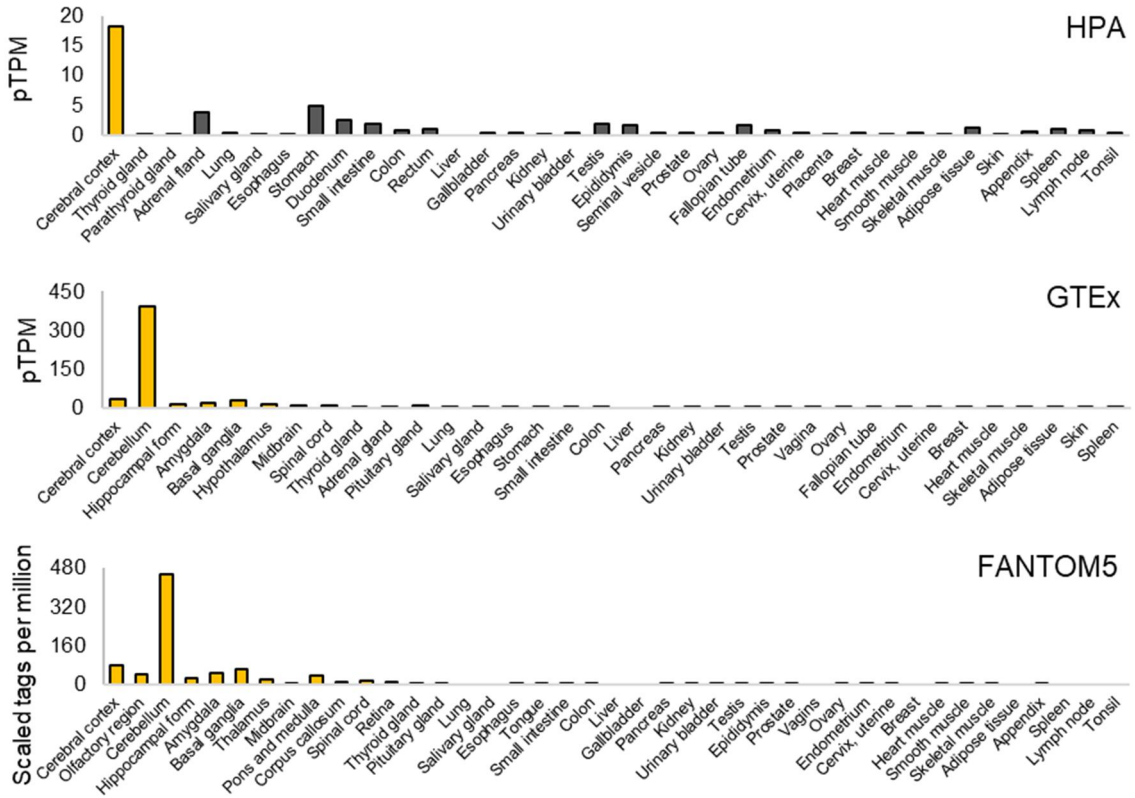


Figure 9. Continued

B

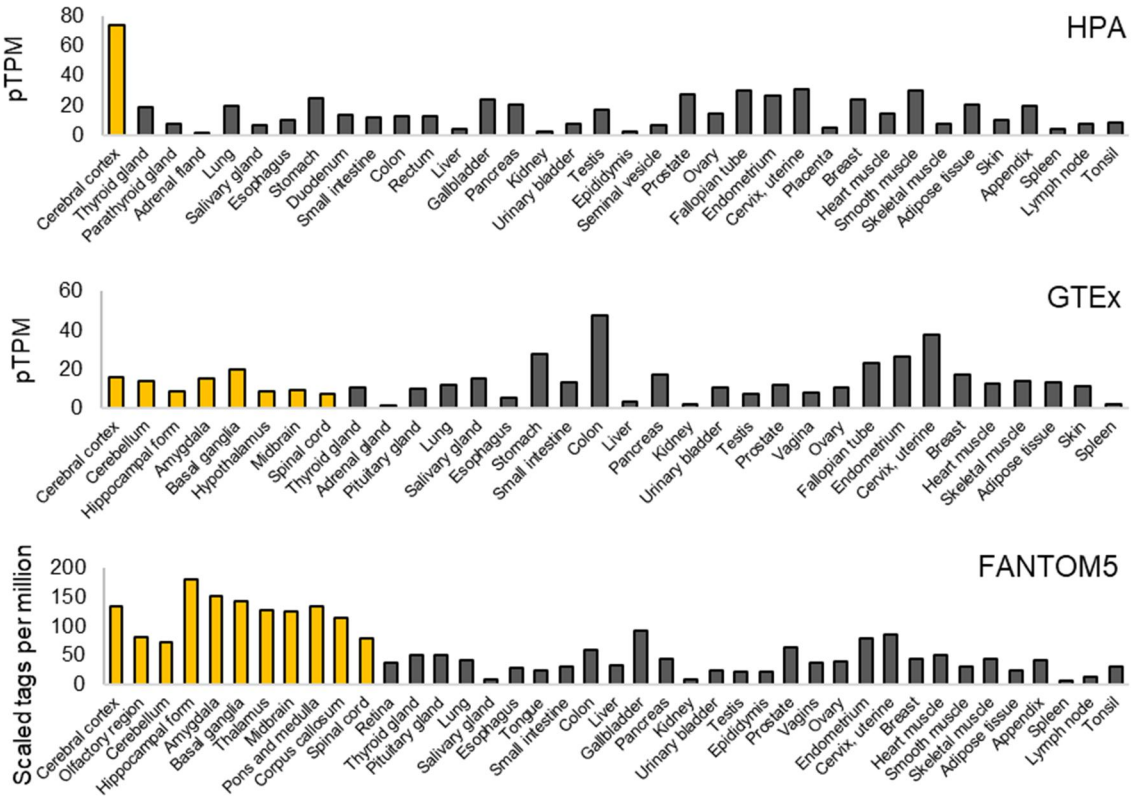


Figure 9. **The RNA expression and distribution of *CACNA1A* (A) and *LRIG1* (B) in brain region (yellow bar) among normal human tissues.** (A) Data from HPA dataset showing highest mRNA expression of the *CACNA1A* gene in the cerebral cortex at 18.2 protein-coding transcripts per million (pTPM). The expression of the *CACNA1A* gene was highest in the cerebellum at 394.1 pTPM in GTEx, 453.4 scaled-tags per million reported by FANTOM5. (B) The expression levels of *LRIG1* were high in the cerebral cortex at 74.1 pTPM in HPA, 16.0 pTPM reported by GTEx. Data from FANTOM5 dataset showing highest mRNA expression of the LRIG1 gene in hippocampal formation at 179.9 scaled-tags per million. (Data available from https://www.proteinatlas.org/ENSG00000141837-CACNA1A/tissue and https://www.proteinatlas.org/ENSG00000144749-LRIG1/tissue)

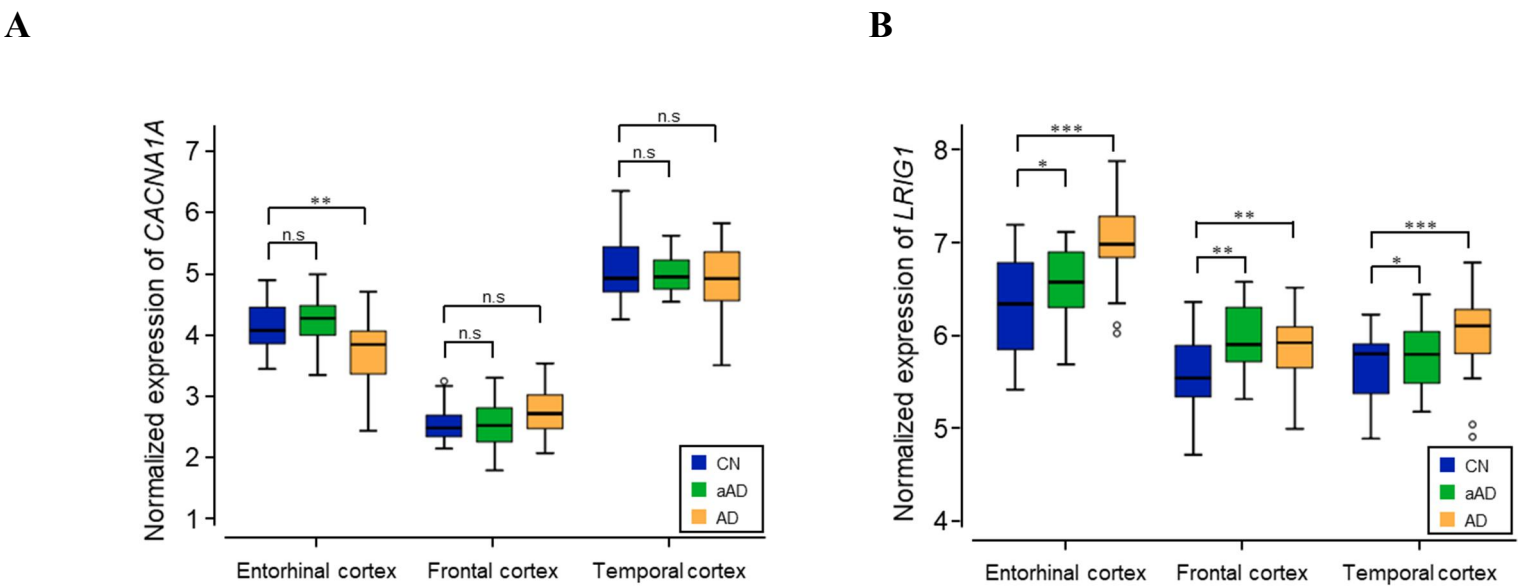


Figure 10. **The differential expression (DE) of *CACNA1A* (A) and *LRIG1* (B) in AD patients' brain.** (A) The *CACNA1A* gene showed significant difference in expression between CN and AD in the entorhinal cortex (p=0.0019). (B) The *LRIG1* gene showed significant differences in expression between the pairs among CN, aAD, and AD in the entorhinal cortex, frontal cortex, and temporal cortex. The significance was labeled above the box plots (*p<0.05, **p<0.01, ***p<0.001). Abbreviations: CN, cognitive normal; AD, Alzheimer's disease dementia; aAD, asymptomatic AD; n.s, not significant

II-4. Discussion

In this study, Genome-wide analysis with a Korean sample was conducted to identify AD-associated loci. The association between AD status and *APOE* locus in chromosome 19 was validated with genome-wide significance. In addition to *APOE*, *SORL1* has also been identified as a potential risk factor for AD with highly significant association in certain East Asian population (Miyashita et al., 2013, Zhang et al., 2017). However, this highly significant association of *SORL1* with AD is greatly influenced by the presence or absence of e4 allele of APOE. To mask the effect of APOE and identify novel candidate loci in East Asians, APOE e4-stratified GWAS was carried out with two independent East Asian samples and identified *CACAN1A* (lead SNP: rs189753894) and *LRIG1* (lead SNP: rs2280575 and rs113765515) among APOE e4 non-carriers. The candidate risk gene, *CACNA1A*, is located on chromosome 19 as in *APOE*. However, *CACNA1A* is far from the *APOE* region (more than 32Mb) and is identified from APOE non-carrier study. For these reasons, it is considered to be unaffected by the *APOE* signal.

From the functional annotation and literature survey, those two genes play an essential role in neurobiological function. *CACNA1A* is the coding gene for the subunit of the calcium channel in neuron. Calcium is not only used a neurotransmitter, but is also used in various parts of the biological system such as cell migration, differentiation, and maintenance. Calcium, reported as a major factor in AD research, is well known for causing neuronal cell death by excessively increasing its concentration in the cytoplasm and mitochondria due to amyloid-beta protein (Demuro et al., 2011, Itkin et al., 2011, Gurkoff et al., 2013). The calcium channel-related genes that have been confirmed to be associated with AD through genomic analysis are *CACNA1C* and *CACNA2D3 (Hollingworth et al., 2012, Tosto et al., 2015)*, and *CACNA1A* has not yet been reported.

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 (Alsina et al., 2016). 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 (Bramham and Messaoudi, 2005, Lu et al., 2013). BDNF showed low gene expression levels in patients with AD (Ng et al., 2019). Several studies have shown that high expression levels of BDNF could slow down cognitive decline in the elderly and in patients with AD (Laske et al., 2011, Lu et al., 2013, Jiao et al., 2016, Buchman et al., 2016).

The novel loci in *CACNA1A* and *LRIG1* genes that were not revealed in large-scale GWASs using Caucasian could be discovered in this study. The allele frequency of rs18973894 in *CACNA1A* is 0.082 in Korean, but 0.007 in European indicate that there is difference in AD associations between East Asian and Caucasian. The AD risk signals that emerged from the meta-analysis with a much larger sample size of Caucasians in the Alzheimer's Disease Sequencing Project (ADSP) appeared to differ considerably from the results of the present study. The 3:66455703 in *LRIG1* showed $P=6.2\times10^{-12}$, and the 19:13395952 in *CACNA1A* showed $P=3.8\times10^{-27}$. Except for the APOE gene, the *SORL1* gene achieved at significance level in each population in East Asian (Miyashita et al., 2013, Zhang et al., 2017). However, its significance depends on whether APOE e4 is present or absent (Jin et al., 2014).

There are limitations to this study that need to be considered. The sample size in this study is relatively smaller than those in other GWASs of AD. Also, this study used only Korean and Japanese populations, not include other East Asians. However, the result of this study was promising due to the high reliability of accurate diagnostic

information based on pathophysiology available in GARD dataset as well as Japanese dataset. Another limitation of this study is the result of imputed data using HRC reference. Although the imputed data may have low accuracy due to lack of Korean references in HRC reference panel, we believe that it facilitated comparison with Japanese data because the advantage of identifying various SNPs from large number of haplotypes.

Previous studies on the function of CACNA1A and LIG1 suggest a role for them in LOAD development, but conclusions on whether each gene influences dementia progression or protects neurons from degeneration require further biological validation through AD model systems. Since the conclusions drawn in this study also acknowledge that they are the result of analyzing two East Asian populations (Korean and Japanese), it is necessary to be cautious in the process of interpreting the results in order to conclude whether the findings are specific to the East Asian population or to the Japanese and Korean populations. Additional replication analysis or meta- or mega-analysis of a larger sample size of other East Asian individuals is strongly recommended in order to conclusively arrive at such an interpretation.

To conclude from the findings made in this work, a portion of previously reported LOAD-associated genes that identified in mostly Caucasian subjects were validated using Koreans. Also, the East Asian specific novel loci were identified through APOE stratified GWAS. The finding in this work might provide an improved understanding of complex genetic mechanisms underlying the development of AD.

PART III

Bootstrap-based genome-wide association studies identify the association of membrane-trafficking pathway genes with Alzheimer's disease

SUMMARY

The onset of Alzheimer's disease (AD) is caused by multiple factors, with genetic factors playing the major role. Although more than 30 AD-related genes have been identified, they are still insufficient to determine the cause of dementia. In this study, a new approach to identify AD-related genes is presented. Bootstrap-based random sampling was applied to GWAS. All of 732 genes from bootstrap-based GWAS were considered as AD-related candidates. A total of 115 genes including 42 genes that had not previously been reported to be associated with AD, were discovered from *Drosophila* scoring. Those genes were validated in biological function through *Drosophila* model of AD. The biological pathways of each gene were determined, and about half of the genes were found to be involved in the membrane trafficking pathway. Of these, 5 genes (*AP3D1*, *CCDC22, DNM3*, *PIK3C3*, and *SH3GL2*) were functionally validated through further functional annotation studies. Significantly, the brain cortex of human AD patients had lower protein levels for these genes than age matched controls. In conclusion, novel risk genes for AD have been identified, suggesting that genes associated with membrane trafficking pathway are likely might be involved in the pathological mechanism of AD.

Ⅲ-1. Introduction

Dementia is a symptom of cognitive decline caused by a variety of diseases, with Alzheimer's disease (AD) being the most common causative disease. The most common early symptom of AD is changes in the brain regions responsible for learning and memory, which typically make it difficult to form new memories and remember recently learned information. Serious memory loss, confusion and disorientation, changes in mood and behaviour, and deeper confusion about events, time, and place are clear features of Alzheimer's disease. In AD patients, pathological changes are observed in the brain: senile plaque, neurofibrillary tangle, and brain atrophy. Senile plaques are produced by the accumulation of beta-amyloid protein on the external surface of neurons, while neurofibrillary tangles are the result of hyperphosphorylated tau protein on the internal surface of neurons. Small accumulations of beta-amyloid, called plaques and oligomers, can contribute to neuronal damage and death by interfering with communication between neurons at the synapse. Tau tangles inside the neuron block the transmission of nutrients and other molecules essential for normal function and neuron survival. Although the entire sequence of events is unclear, beta-amyloid begins to accumulate prior to abnormal tau, and increased accumulation of beta-amyloid is associated with subsequent increases in tau (Bloom, 2014). In AD, the accumulation of beta-amyloid and phosphorylation of tau protein causes the death of neurons throughout the brain, severing the neuronal network and reducing many brain regions. Hippocampus, entorhinal cortex, subcortical structures, and white matter were observed in a large part of the brain, and the volume of the brain was greatly reduced (Pini et al., 2016).

Genetic factors are consistently regarded as important predictors in understanding the mechanisms of AD. Among the genetic factors associated with AD, the most critical and consistently reported gene is Apolipoprotein E (APOE), which explains the largest portion of the disease (Strittmatter et al., 1993). Except

for APOE, a number of AD-related genes, such as *CD33, ABCA7, BIN1, CLU, EPHA1* and *SORL1*, have been identified by genome-wide association study (GWAS) (Harold et al., 2009, Hollingworth et al., 2011, Miyashita et al., 2013). In 2013, 11 new AD genetic variants were discovered through a meta-analysis using 74,046 individuals (Lambert et al., 2013). The International Genomics of Alzheimer's Project (IGAP) in 2019, 24 AD-related genes were discovered using 94,437 individuals (Kunkle et al., 2019). Most recently, using more than a 1 million people from all cohorts combined, 39 genes were confirmed to have associations with AD (Wightman et al., 2021). The discovered genes, however, were insufficient to explain the pathogenesis and mechanisms of AD. In addition, many genes have not been found to be involved in biological function of AD. Also, many of the SNPs, which were identified the association with AD, are located in introns or intergenic regions which function is not precisely known (Freedman et al., 2011, Blattler et al., 2014).

Various statistical attempts for GWAS associated with diseases have led to discover the novel genes (Faye et al., 2011, Vélez et al., 2013, Arbet et al., 2017). Bootstrap sampling, one of the sampling techniques, is a technique to extract a large number of smaller sample size groups. Bootstrapping analysis was useful when the sample size is small. There are several reports of GWAS results using the bootstrap sampling, but the analysis results using the Korean genome are insufficient.

The genes discovered from GWAS were constantly required additional research works to validated biological function. The study of functional genomics using small genetic models such as *Drosophila* is a good counterpoint to overcome the limitations of GWAS. Drosophila is a well characterized genetic model for testing a large number of genes within a short time. Also, Drosophila have genes with capabilities that correspond to 75% of human genes (Bier, 2005). The efficiency of the Drosophila-based GWAS has been validated through several reports (Shulman et al., 2011, Wangler et al., 2017).

In this study, bootstrap-based GWAS using Koreans was used to identify the novel AD-related genes. In addition, the function of the discovered candidate genes was validated through scoring using the *Drosophila* model. Functional analysis with public databases was used for annotate the biological function and pathway. As a result, five genes (*AP3D1*, *CCDC22*, *DNM3*, *SH3GL2*, and *PIK3C3*) were found to be novel AD-related genes. These genes were associated with membrane trafficking pathway and were identified as regulating the amyloid accumulation.

III-2. Methods

III-2-1. Study participants

This study utilized data from participants aged over 60 years from the GARD. Out of 14,524 participants, 2,291 (1,119 AD dementia and 1,172 cognitive normal) subjects were chosen to maximize contrast between cases and controls for the GWAS. The cases had an age distribution of 60 to 97 years and met the NINCDS-ADRDA for AD. Those with amyloid negative diagnosis by PET imaging were excluded. The control group consisted of individuals whose age distribution ranged from 70 to 90 years and who were cognitively normal or did not exhibit neurological, psychological, or pathological symptoms. For the expansion of hospital-based samples, blood or DNA with clinical information was received from 13 institutions, including Seoul National University Hospital, Inha University Hospital, Gyeongbuk University Hospital, Dong-A University Hospital, and Pusan National University Hospital. The study protocol was approved by the Institutional Review Board of Chosun University Hospital, Korea (CHOSUN 2013-12-018-070).

III-2-2. SNP genotyping

Genomic DNA from GARD cohort was extracted from peripheral blood leukocytes that were isolated from whole blood collected in EDTA tube. Of the total

subjects, 5,570 blood samples were finally used for genomic analysis, except for diagnostic uncertainty, mixed-up, and inadequate DNA concentrations among all subjects. The samples were genotyped using the KNIH Biobank Array as part of the discovery sample. Affymetrix Power Tools (APT) were used to process all CEL files. Dish QC (DQC) values were generated and used to remove samples with DQC<0.82. Low-quality markers selected by the Ministry of Health to improve marker quality control have been removed. Also, low-quality markers were removed by SNPolisher to pass through a better sample QC procedure. The 4,391 samples used KNIH Biobank array v1.0, while 1,179 samples used KNIH Biobank array v1.1 which is a more supplemented by Koreans. The number of samples analyzed for bootstrap-based GWAS was expanded by integrating both versions and extracting common SNPs.

III-2-3. Quality control

Quality control was performed under the same conditions for each chip version using the PLINK v1.90 package (Chang et al., 2015). SNPs with genotype call rate (GCR) < 95%, a minor allele frequency (MAF) <0.01, or significant deviation from the Hardy-Weinberg equilibrium (HWE) ($P<10^{-6}$) were excluded. Samples with individual call rate < 95%, gender inconsistency between reported and analysis of X-chromosome SNPs, extremely low or high genome-wide heterozygosity (±3 S.D from the mean). After performing QC for SNPs and samples, 4,011 samples and 518,326 SNPs were passed from Korean chip v1.0 and 1,117 samples and 577,694 SNPs were passed from Korean chip v1.1. For bootstrap-based GWAS, totally 2,291 subjects (1,172 controls and 1,119 AD cases) were selected.

III-2-4. Imputation

SNP genotypes for bootstrap GWAS were imputed separately using reference haplotypes pre-phased in the HRC (Haplotype Reference Consortium)

panel version 1.1 (McCarthy et al., 2016). Eagle (version 2.3) was used for phasing, and Minimac4 was used for imputation. The low-quality (info score <0.5) imputed SNPs were removed. After imputation, 35,685,761 variants were obtained by combining the data from Korean discovery dataset.

III-2-5. Bootstrap-based sampling for GWAS

The following resampling approach was used for generating a large number of candidate genes. From the original sample with size *n*, *B* bootstrap samples with varying size *k* are generated. The model took $B = 100$ and k be a randomly chosen integer within $0.5n < k < n$. For each bootstrap sample, perform a genome-wide analysis using logistic regression and adjust for age and sex. The candidate gene will report and include SNPs (or genes) with a P-value of less than 5×10^{-5}. These genes were simply combined with the results, based on whether or not they showed suggestive significance in multiple analysis. SNPs with higher or lower Odd ratios, not p-values, were selected as the final candidate. These methods aim to capture as many candidate genes as possible, and therefore false-positives are likely to be inherent.

III-2-6. Public database filtering

To detect potential AD risk genes, previously reported genetic functions were annotated using publicly available databases. Candidate genes for which gene expression in brain regions has not been reported in the Genotype-Tissue Expression (GTEx, release v8) (https://gtexportal.org/home/) (Consortium et al., 2015) and Human Protein Atlas (HPA) (http://www.proteinatlas.org/) (Uhlén et al., 2005) , and Functional Annotation of Mammalian Genomes 5 (FANTOM5) (https://fantom.gsc.riken.jp/5/) (Lizio et al., 2015) were excluded.

III-2-7. Functional screening of candidate genes using *Drosophila*

The rough phenotype caused by amyloid accumulation in Drosophila eyes was the criteria to be scored. Inhibition of candidate gene expression through RNAi can cause conversion of Drosophila eye traits or prevent them from changing. Genes that further aggravate the phenotype were scored positive, and gene that are protected were scored negative. The *Drosophila* experiment was supported by Professor Kyoung Sang Cho's team at Konkuk University.

III-3. Results

III-3-1. Bootstrap-based GWAS discovered AD-related genes

For the study sample, 2,291 participants (1,119 AD and 1,172 controls) from the Gwangju Alzheimer's Related Dementias (GARD) were included. To provide a better opportunity to identify candidate genes associated with AD, a bootstrap-based GWAS was performed. In particular, GWAS was performed on 100 bootstrap samples originally extracted from the samples, and a genome-wide suggestive level of 5×10^{-5} was applied to define the candidate gene list. These methods aim to capture as many candidate genes as possible, and therefore false-positives are likely to be inherent. As a result of the GWAS using bootstrapping, 732 genes satisfying the genome-wide suggestive level were selected as the candidates.

To evaluate the performance of bootstrap GWAS, International Genomics of Alzheimer's Project (IGAP) and AlzGene from public database were used to compare the number of genes. Overlapping genes with the standard GWAS was only 1, while bootstrap GWAS found 52 genes overlapped with AlzGene. When compared to the genes reported in IGAP, more genes are identified in the bootstrap GWAS than in the standard GWAS at levels with $P<0.05$, $P<0.01$, $P<1\times10^{-4}$, $P<1\times10^{-5}$. Moreover, the previously reported as AD associated genes (such as *ABCA7*, *PICALM*, and *SORL1*) appear in the bootstrap GWAS results. To determine

novel AD-related genes, genes that have not been reported to be associated with AD were selected as candidate groups.

Table 5. Summary of demographic information for bootstrap GWAS

	CN	AD
All samples, n	1,172	1,119
Female, n (%)	661 (56.4)	715 (63.9)
Age at exam, m (s.d)	76.03 (8.9)	74.60 (8.9)
APOE e4 non-carrier, n	976	621
APOE e4 carrier	196	498

Abbreviations: CN, cognitive normal; AD, Alzheimer's disease; s.d, standard deviation.

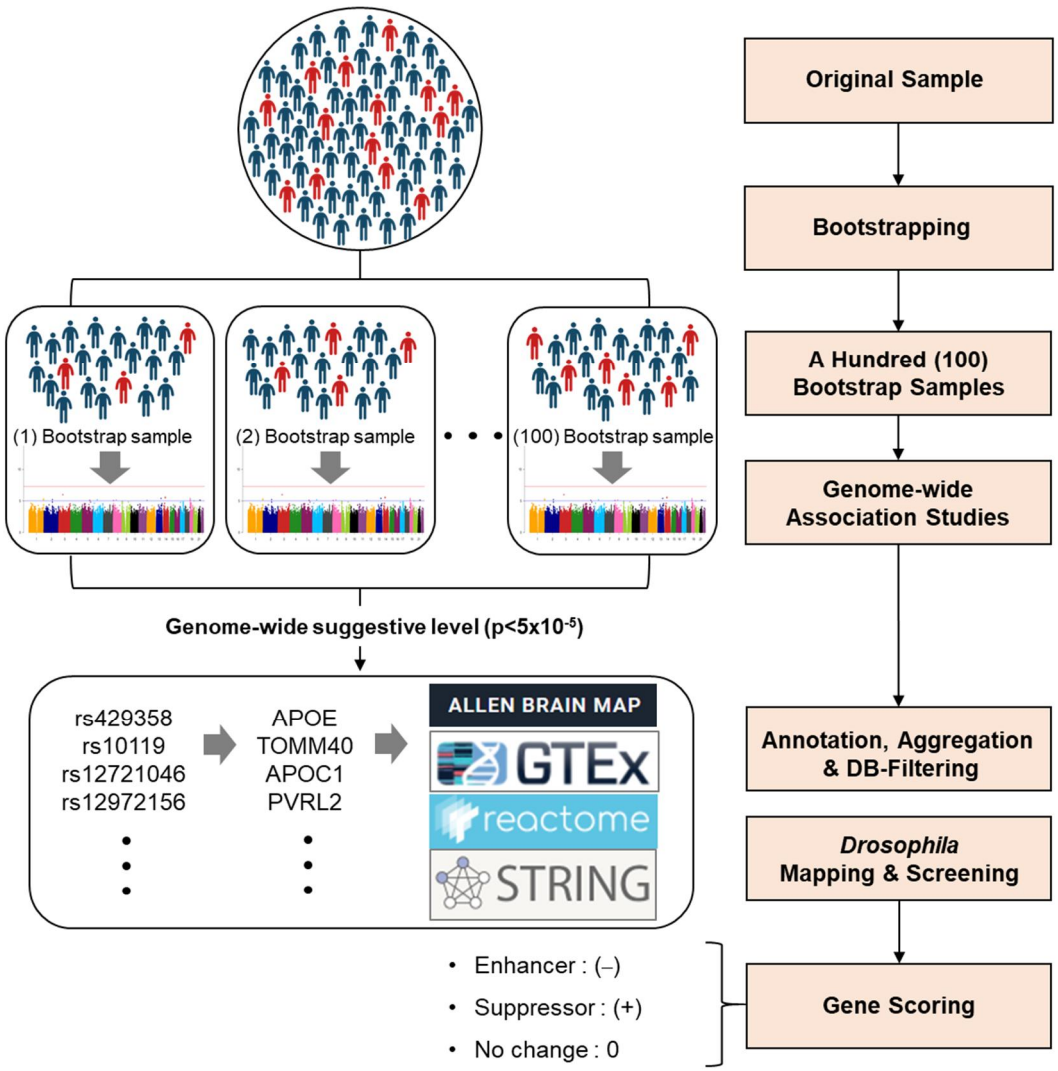


Figure 11. Overall process of bootstrapping GWAS and scoring genes.

III-3-2. Functional genomic screening of the AD-associated genes using the *Drosophila* AD model

To examine the biological functions for the selected as candidate genes through bootstrap GWAS, *Drosophila* AD model was applied for functional genomic screening. The model, in which the human beta-amyloid 42 transgene is expressed in the eye, induced a rough phenotype. The 732 human genes were matched with 507 fly genes. To characterize the function of candidate genes, RNAi-based gene repression was induced in *Drosophila* eyes in which beta-amyloid 42 was being expressed. Then, the phenotypes of eyes expressing only amyloid and eyes expressing RNAi along with amyloid were compared and scored. The results showed that 85 genes aggravated the eye phenotype and 18 genes suppressed the eye phenotype. 404 genes caused no change in the phenotype. A total of 115 human genes were used in further functional studies.

III-3-3. Most of the candidate genes were involved in membrane-trafficking pathways

To develop an insight into the biological pathways involved in Aβ pathology, A total number of 115 human genes (which were annotated with 103 *Drosophila* genes) was selected to analyze the biological function. Through literature and public database search (gene ontology and KEGG pathway), 42 out of 115 human genes were found to be associated with membrane trafficking pathway. The membrane trafficking, which involves the transport of proteins and neurotransmitters during neuronal development, can lead to neurodegenerative disorders when defective (Wang et al., 2013, Winkle and Gupton, 2016). From the 42 genes, 5 genes (*AP3D1*, *CCDC22*, *DNM3*, *SH3GL2*, and *PIK3C3*) were finally selected. These genes were the ones that have not been reported in association with AD. Also, restricting the expression of these genes in the *Drosophila* eye increased amyloid-beta toxicity.

This suggests that defect of membrane trafficking is associated with enhanced intensified amyloid toxicity to cells.

Table 6. Overlap with publicly available AD-associated genes

	N	AlzGene	IGAP ($P<0.05$)	IGAP ($P<0.01$)	IGAP ($P<1\times10^{-4}$)	IGAP ($P<1\times10^{-5}$)
Standard GWAS	92	1	60	39	5	2
Bootstrap GWAS	732	52	521	289	21	7

**APOE*, *APOC1*, *APOC1P1*, *TOMM40*, and *PVRL2* are excluded.

Table 7. List of AD-related candidate genes involved in amyloid pathology.

Chr	SNP	BP	Gene code	OR
1	AX-40230529	95692909	ENST00000303635	1.92
1	AX-16677431	15098897	ENST00000355305	2.656
1	AX-11131178	18373797	ENST00000304685	2.108
1	AX-31891363	3836884	ENST00000371079	1.903
1	AX-16441570	67792359	ENST00000357650	1.337
1	AX-113686088	52246936	ENST00000370855	4.748
1	AX-113207361	2127493	ENST00000307340	2.269
2	AX-83215840	112274613	ENST00000312428	2.065
2	AX-114350156	106695625	ENST00000307845	1.638
2	AX-114334954	117586284	ENST00000332345	1.671
2	AX-113334323	61727485	ENST00000263816	2.821
2	AX-11180835	22044324	ENST00000302217	1.913
2	AX-114112618	96047736	ENST00000252711	1.852
2	AX-33089401	5615437	ENST00000234111	1.464
2	AX-113827487	7758798	ENST00000294964	1.602
2	AX-33537227	60618452	ENST00000455896	0.549
2	AX-82984255	123271680	ENST00000171887	1.803
3	AX-11284410	11073961	ENST00000283290	1.156
3	AX-114164301	155366240	ENST00000263665	2.442
3	AX-114300364	74488051	ENST00000357716	2.634
3	AX-12609846	146937656	ENST00000263754	1.765
3	AX-34205077	76393989	ENST00000296252	1.387
3	AX-14215539	149488870	ENST00000414318	2.17
3	AX-107719170	68265576	ENST00000284483	2.74
3	AX-113170079	31509043	ENST00000296328	1.414
4	AX-34819019	169359519	ENST00000330055	2.026
4	AX-14801403	4372194	ENST00000264568	1.859
4	AX-113657001	1641830	ENST00000260184	1.865
4	AX-113197539	21845330	ENST00000506970	2.757
4	AX-12636562	196852887	ENST00000327908	1.224
4	AX-14678063	11795157	ENST00000295971	1.99

4	AX-114293091	171843791	ENST00000342295	2.277
4	AX-11569982	41700214	ENST00000407365	2.284
5	AX-14929144	23534459	ENST00000286301	2.307
5	AX-14926141	73569926	ENST00000314512	2.603
5	AX-114176843	45680041	ENST00000512467	2.071
5	AX-11396541	133748644	ENST00000199814	0.932
5	AX-35299969	123103828	ENST00000506237	1.885
6	AX-113144694	123195137	ENST00000343245	2.573
6	AX-15355440	29050369	ENST00000376185	1.72
6	AX-67483945	60298560	ENST00000355167	1.524
6	AX-67450439	86644999	ENST00000356535	1.78
6	AX-35760071	34108564	ENST00000374177	1.242
6	AX-41961279	7226581	ENST00000374316	1.394
6	AX-15413156	8846649	ENST00000281156	1.883
6	AX-15374416	10498095	ENST00000230418	1.406
6	AX-35564077	35551427	ENST00000275230	1.46
6	AX-113815388	147907579	ENST00000333572	2.058
7	AX-36111861	15182889	ENST00000321736	1.981
7	AX-113673676	26742972	ENST00000361727	3.024
7	AX-114165630	33630472	ENST00000328843	2.85
7	AX-113633109	20136485	ENST00000265753	1.43
7	AX-15514755	143571950	ENST00000340010	1.299
7	AX-36281705	63152630	ENST00000421025	1.455
8	AX-36571027	136614486	ENST00000306349	1.438
8	AX-36543833	94374603	ENST00000421627	2.766
8	AX-15882905	109248169	ENST00000355849	2.25
8	AX-36784471	185235226	ENST00000302190	1.7
8	AX-114334742	110786290	ENST00000382080	2.593
8	AX-113149199	170195241	ENST00000419617	2.137
9	AX-36945551	30355102	ENST00000380548	2.389
9	AX-36855517	102342511	ENST00000349780	1.831
9	AX-113711505	116637083	ENST00000380607	4.38
9	AX-113198274	50555983	ENST00000374994	1.472
9	AX-114163314	240946373	ENST00000376552	1.765
10	AX-83342145	33494039	ENST00000378211	1.627

10	AX-114306172	88481623	ENST00000303745	3.353
10	AX-114136719	10586660	ENST00000315238	1.36
10	AX-37457727	40561246	ENST00000379261	1.784
10	AX-16384484	57311728	ENST00000373744	2.725
10	AX-12435167	39482779	ENST00000369061	2.093
10	AX-11149560	42325918	ENST00000260731	1.83
10	AX-114112885	46686080	ENST00000373957	2.338
10	AX-114317272	45630779	ENST00000361972	2.011
11	AX-113874428	43073085	ENST00000007633	1.705
11	AX-83524439	23494119	ENST00000301774	1.433
11	AX-84943959	150087887	ENST00000227520	1.925
11	AX-11196920	40516682	ENST00000379554	1.912
11	AX-29861657	26336442	ENST00000335953	1.874
12	AX-11432354	183617094	ENST00000301190	1.362
12	AX-113196754	64581937	ENST00000381340	2.363
12	AX-107901847	55186091	ENST00000281928	1.966
12	AX-113649407	97551919	ENST00000344941	1.89
13	AX-114311195	43957510	ENST00000376887	2.209
13	AX-113706959	56761069	ENST00000282397	2.137
14	AX-31338883	7427402	ENST00000341321	2.453
14	AX-113146982	15006623	ENST00000251091	1.7
14	AX-31500777	17616366	ENST00000555005	2.443
15	AX-107778825	16217238	ENST00000560857	1.583
15	AX-12987420	107441324	ENST00000355254	1.966
15	AX-31600117	50523757	ENST00000260404	1.382
15	AX-12921946	134435705	ENST00000267842	1.496
16	AX-113692167	101376836	ENST00000307431	2.211
16	AX-54654994	75906959	ENST00000449606	1.482
16	AX-12429419	220483371	ENST00000320241	1.563
17	AX-106717449	99556509	ENST00000262442	2.863
18	AX-113202501	43689201	ENST00000315677	2.361
18	AX-114130267	32551388	ENST00000262039	2.045
18	AX-113864205	33193617	ENST00000587834	1.329
18	AX-113627154	38228570	ENST00000585404	4.416
19	AX-32656553	17962362	ENST00000345016	1.804

19	AX-32713813	101916315	ENST00000262626	1.843
19	AX-32751013	82169138	ENST00000221462	1.224
19	AX-40480671	170870307	ENST00000269829	2.046
20	AX-113183526	218713365	ENST00000317304	2.115
20	AX-11706390	90966179	ENST00000372733	0.925
20	AX-113710195	196141815	ENST00000372801	3.144
20	AX-114109228	216457245	ENST00000216923	2.359
21	AX-33281443	169987027	ENST00000400454	3.365
21	AX-113644983	113879295	ENST00000361371	1.496
21	AX-113621101	45669825	ENST00000331343	1.485
22	AX-113643142	50797025	ENST00000262794	2.044
22	AX-40854533	58772579	ENST00000263116	1.15
22	AX-113912696	63749830	ENST00000332840	2.037

Abbreviations: Chr, chromosome; BP, base-pair position; SNP, single nucleotide polymorphism; OR, odds ratio.

III-4. Discussion

In this study, bootstrap-based GWAS was conducted using Korean sample to discover the novel genes associated with AD. Moreover, the *Drosophila* model served to filter out AD-related genes and validate their functions. Bootstrap-based GWASs can identify more AD-responsible genes than standard GWAS. Compared to the standard GWAS, which exclusively matched only 65% of the genes assumed, the bootstrap-based GWAS matched 71.2% of the genes from IGAP ($P<0.05$). Out of 115 genes from *Drosophila* scoring, 42 genes were found to be associated with membrane trafficking pathways. This suggests that the membrane trafficking pathway is associated with amyloid accumulation and toxicity. From the functional analysis using public databases, five genes (*AP3D1*, *CCDC22*, *DNM3*, *SH3GL2*, and *PIK3C3*) were found to be novel AD-related genes. These genes were associated with membrane trafficking pathway and were identified as regulating the amyloid accumulation.

The membrane trafficking pathway is a very diverse and complex process involved in the transport of substances into and out of cells. Presenilin1 and tau are the well-known trafficking proteins in AD researches (Howell et al., 2006, Rajendran and Annaert, 2012). Abnormal cleavage of amyloid precursor protein (APP) by beta- and gamma-secretases causes trafficking defection and accumulating the beta-amyloid, which leads to neurodegeneration (Uemura et al., 2004, Rajendran and Annaert, 2012). The beta-secretase activity is induced by a membrane protein called beta-site APP-cleaving enzyme 1 (BACE1), and the gamma-secretase is mediated by a membrane protein complex called presenilins (PSENs) (Epis et al., 2012). Unfortunately, there have been insufficient researches explaining the membrane trafficking mechanisms for other genes or proteins associated with AD.

From the functional annotation and literature survey, those two genes play an essential role in neurobiological function. According to GeneCards

(http://www.genecards.org), Adaptor Related Protein Complex 3 Subunit Delta 1 (AP3D1) protein is associated with the golgi region and implicated in trafficking of neurotransmitter vesicles. Also, the adaptor related protein (AP) 3 plays a major role in the transport of molecules in neurons, such as in the sorting the vesicles and axonal targeting of transporters (Guardia et al., 2018).

Coiled-Coil Domain Containing 22 (CCDC22) protein is involved in regulation of NF-kappa-B signaling (Starokadomskyy et al., 2013). The NF-kappa-B signaling plays role in immune response, cell apoptosis, proliferation, and differentiation. This complex, which is also involved in ubiquitination, gene transcription, and other biological processes, plays a major role in the transport of proteins in cytoplasm. Moreover, NF-kappa-B also acts as a regulator of hippocampal synaptic plasticity (Albensi and Mattson, 2000).

Dynamin 3 (DNM3) protein is subfamily of GTP-binding proteins that involved in microtubule association and vesicle transport. The interaction of the microtubule associated protein tau and dynamin has been reported previously. Induced expression of tau, which is highly associated with AD, in cortical neurons showed decreasing the level of dynamin 1 and caused defects in endocytosis (Xie et al., 2019). Although no association between dynamin 3 and AD has been reported, dynamin 1 and dynamin 2, which has a similar function, has been reported to be highly associated with AD (Aidaralieva et al., 2008, Zhu et al., 2012, Oliver and Reddy, 2019).

SH3 Domain Containing GRB2 Like 2, Endophilin A1 (SH3GL2) protein is a member of endophilin A, which is increased in AD brain (Ren et al., 2008). Increased endophilin A1 at synaptic terminals led to synaptic alterations from the AD model (Yu et al., 2018). However, these studies were validated in animal models, not in human AD patients, and the mechanisms may differ from human AD. Endophilin A1 is also associated with morphogenesis, especially formation of

dendritic spines, which is known to affect synaptic plasticity and potentiation (Yang et al., 2015, Yang et al., 2018).

Phosphatidylinositol 3-Kinase Catalytic Subunit Type 3 (PIK3C3), also known as VPS34, is a protein that make complex with other membrane proteins, taking part in a variety of membrane trafficking processes in neurons. The defects in PIK3C3 can lead to loss of synapses, neurodegeneration, and dysregulation of intracellular amyloid precursors (Wang et al., 2011, Morel et al., 2013, Miranda and Di Paolo, 2018). However, the AD association of the gene through genomic analysis has not yet been reported.

In conclusion, this study used a bootstrap-based GWAS to discover 115 AD-related genes in Koreans. Moreover, the discovered genes have been screening and functionally validated using *Drosophila* AD model. Considering that 42 genes are associated with membrane trafficking pathways, the results suggest that membrane trafficking and amyloid accumulation are highly correlated in AD pathology. Moreover, 5 candidate genes (*AP3D1*, *CCDC22*, *DNM3*, *SH3GL2*, and *PIK3C3*) were investigated for their biological functions, establishing their importance for beta-amyloid toxicity, neuronal degeneration, and synaptic plasticity. The association of these proteins with AD was confirmed indirectly through their association with amyloid protein. This study presented a bootstrap sampling for GWAS in Korean population and a process of functional validation using *Drosophila* AD model. Identification of novel AD-related genes played role in membrane trafficking pathway will contribute to understanding of the genetic mechanisms of AD.

PART IV

Protective effect of APOE e2 for late-onset Alzheimer's disease in Korean

SUMMARY

The effect of Apolipoprotein E (APOE) gene region may explain ethnic differences in the association of Alzheimer's disease (AD) with e4. More than one copy of APOE e4 allele increases the risk of AD, while more than one copy of APOE e2 allele decrease the risk of AD. The effects of APOE e2 homozygotes were not well defined because of its low frequency worldwide. In this study, the effects of APOE e2 were confirmed in Korean population. For contrast odds ratios, the association analysis was performed separately for the group with clinical diagnosis and the group with neuro-pathological diagnosis using amyloid PET images. The association analysis confirmed that the risk of AD was 0.39 times lower in the group with APOE e2 than in the group with APOE e3/e3. And those protective effects were more evident in the group with pathologically confirmed group. Through survival analysis, the age of onset of AD was delayed by 3 years in the group with e2-carrier compared to group with e3/e3. The degree of amyloid-beta accumulation with advancing age showed that those with at least one e2 had no or less amyloid accumulation than those without. In this study, the effect of APOE e2 associated with AD in Korean was conducted using GARD cohort. Also, this study shows that different patterns of amyloid beta accumulation by APOE allele types according to age.

IV-1. Introduction

The most genetically established Alzheimer's disease (AD) risk factor is the e4 allele of the Apolipoprotein E (APOE) gene on chromosome 19. *APOE* is a gene that codes for proteins involved in fat metabolism, such as cholesterol transport in neurons (Martins et al., 2006). The human *APOE* gene has three isotypes e2, e3, and e4 depending on the combination of two SNPs (rs429358 and rs7412), with frequencies of 8.4%, 77.9%, and 13.7%, respectively (Farrer et al., 1997). Since human genome is diploid, *APOE* allele can be inherited from each parent, resulting in six allele types: e2/e2, e2/e3, e2/e4, e3/e3, e3/e4 and e4/e4.

The frequency of the e4 allele is greatly increased in AD patients. Having one copy of the e4 allele of APOE increases the risk of AD 4-fold, and having two copies increases it more than 10-fold (Farrer et al., 1997). The APOE e2 has the lowest frequency in populations worldwide, especially in patients with AD. The e2 allele of APOE provides a protective effect against AD (Corder et al., 1994). When compared to the APOE e3/e3 population, which had the highest frequency, the population with APOE e2 had the lowest risk of AD and the highest mean age (Corder et al., 1994, Farrer et al., 1997). However, the mechanism underlying the protective effect of APOE ε2 against AD remains unclear.

According to a previously reported study, APOE e2 is associated with reduced amyloid deposition in the brain (Xu et al., 2012, Kanekiyo et al., 2014), which at least suggests that APOE e2 reduces the risk of AD. Moreover, APOE e2 prevents from cognitive impairment for subjects with higher beta-amyloid levels in the brain (Kim et al., 2017, Sweigart et al., 2021). The APOE e2 has also been associated with longevity (Vélez et al., 2016, Shinohara et al., 2020, Kuo et al., 2020).

Recently, the Alzheimer's Disease Genetics Consortium (ADGC) reported the protective effect of the APOE e2 genotype using over 20,000 samples (Reiman et al., 2020). For the analysis, neuro-pathologically diagnosed groups by autopsy

were used to ensure diagnostic accuracy. Analyses using a large sample were sufficient to establish an APOE effect, there was no evidence for difference between ethnic groups.

In this study, the genetic information of Koreans, a homogeneous population that maintains a well-defined genetic background with a high prevalence of AD among the elderly over 60 years old, was used to determine the effects of APOE e2. To confirm the powerful contrasting protective effects of APOE e2, subjects were divided into suspected group and neuro-pathologically confirmed group. To confirm the protective effect of APOE in Koreans, age of onset and amyloid accumulation were compared by APOE genotype.

Ⅳ-2. Methods

Ⅳ-2-1. Study participants

For the study, data from participants aged over 60 years from the GARD cohort at Chosun University in Gwangju, were used. Out of 14,524 participants, 3,924 (1,225 AD dementia and 2,699 cognitive normal) participants were chosen to maximize contrast between cases and controls. The cases had an age distribution of 60 to 97 years and met the NINCDS-ADRDA for AD. Those with amyloid negative diagnosis by PET imaging were excluded. The control group consisted of individuals whose age distribution ranged from 70 to 90 years and who were cognitively normal or did not exhibit neurological, psychological, or pathological symptoms. For the expansion of hospital-based samples, blood or DNA with clinical information was received from 13 institutions, including Seoul National University Hospital, Inha University Hospital, Gyeongbuk University Hospital, Dong-A University Hospital, and Pusan National University Hospital. The study protocol was approved by the Institutional Review Board of Chosun University Hospital, Korea (CHOSUN 2013-12-018-070). To confirm the powerful contrasting protective effects of APOE e2,

subjects were divided into suspected group and neuro-pathologically confirmed group. The suspected group diagnosed through neurological examination without amyloid PET imaging and the neuro-pathologically confirmed group diagnosed with amyloid PET.

Table 8. Summary of demographic information for each group.

	Suspected group		Pathologically confirmed group	
	Control	Dementia	Control	ADD
N	1840	770	859	366
Sex, F (%)	1207 (65.6)	515 (66.9)	477 (55.6)	210 (57.4)
Age, M (SD)	72.78 (5.9)	74.92 (6.9)	70.71 (5.7)	73.67 (6.7)
Amyloid PET	-	-	Negative	Positive
SUVR, M (SD)	-	-	1.0125 (0.08)	1.3571 (0.15)

Suspected group, neurologically diagnosed group without amyloid PET; Confirmed group, pathologically diagnosed group with amyloid PET.

Abbreviations: CN, Cognitive normal; NC, Normal control; ADD, Alzheimer's disease dementia

IV-2-2. SNP genotyping and quality control

Genomic DNA from GARD cohort was extracted from peripheral blood leukocytes that were isolated from whole blood collected in EDTA tube. Of the total subjects, 5,570 blood samples were finally used for genomic analysis, except for diagnostic uncertainty, mixed-up, and inadequate DNA concentrations among all subjects. The samples were genotyped using the KNIH Biobank Array as part of the discovery sample. Affymetrix Power Tools (APT) were used to process all CEL files. Dish QC (DQC) values were generated and used to remove samples with DQC<0.82. Low-quality markers selected by the Ministry of Health to improve marker quality control have been removed. Also, low-quality markers were removed by SNPolisher to pass through a better sample QC procedure. The 4,391 samples used KNIH Biobank array v1.0, while 1,179 samples used KNIH Biobank array v1.1 which is a more supplemented by Koreans. The number of samples analyzed for discovery study was expanded by integrating both versions and extracting common SNPs. Quality control was performed under the same conditions for each chip version using the PLINK v1.90 package (Chang et al., 2015). Samples with individual call rate < 95%, gender inconsistency between reported and analysis of X-chromosome SNPs, extremely low or high genome-wide heterozygosity (±3 S.D from the mean).

IV-2-3. Positron emission tomography (PET) image processing

A total of 1,225 participants from GARD had PET scans at the Chosun University Hospital. A Discovery ST (General Electric Medical Systems) PET-CT scanner was used to acquire PET images 90 minutes after injection of approximately 300 MBq18F-Florbetaben (18FBB) intravenously. Each of the subjects' MRT1-weighted images was processed with FreeSurfer (version 5.3.0) using an optimized automated processing flow to reproduce a 3D cortical surface model (Dale et al., 1999, Fischl et al., 1999, Fischl and Dale, 2000).

IV-2-4. Standardized uptake value ratio (SUVR) calculation

All PET images were processed in SPM12 toolbox which is implemented in MATLAB (R2018a, Mathworks, Natick, MA, USA). At the beginning, each T1 weighted MR image of the same subject taken within 6 months of the same time or interval time was co-registered with the F-18 PET image. The transformation parameters were then estimated for the spatial normalization of each T1-weighted image in the standard MNI template. In addition, the PET images were spatially normalized as standard MNI templates using the parameters. FreeSurfer (version 6.0) was used to apply the segmentation of each structural MR image and extract the standard value uptake ratios (SUVR) from 68 predefined regions of interest (Palmqvist et al., 2017). SUVRs for the frontal lobe, lateral parietal, posterior cingulate, anterior cingulate, and lateral temporal regions were used to generate the global cortical retention mean SUVRs for the whole cerebellum (Maass et al., 2017).

IV-2-5. Statistical analysis

To estimate the ORs of AD associated with each APOE allele, the APOE genotype of interest is coded as 1, the reference genotype is coded as 0 (APOEe2/e2, APOEe3/e3 or APOEe4/e4) and the other genotype is coded as missing. The ORs and 95% CIs for each APOE genotype were calculated compared to APOE e3/e3 genotype. A genotype association test was performed for each APOE allele genotype with the APOE e3/e3 used for AD with or without covariate adjustment for age and sex. For estimations, ORs were calculated in logistic regression using a generalized linear model. For allelic dose evaluation of APOE e2, the e3/e3, e2/e3, and e2/e2 were coded as 0,1, and 2, respectively. For allelic dose evaluation of APOE e4, the e3/e3, e3/e4, and e4/e4 were coded as 0,1, and 2, respectively. The Allelic association tests for the APOE e2 and APOE e4 were performed under a logistic regression analysis with or without covariate adjustment for age and sex using the GLM for

AD. Data from the neuro-pathologically confirmed group were used to generate the Kaplan-Meier curves for each APOE genotype. The curve represents the proportion of neuro-pathologically confirmed cases and controls who are free or onset for AD as they age. Best fit lines of correlation between age and SUVR were estimated using linear regression.

Ⅳ-3. Results

Ⅳ-3-1. Neuropathologically confirmed and unconfirmed groups

The number of Alzheimer's dementia cases and cognitively unimpaired controls for each APOE genotype in GARD cohort clinically characterized and neuropathologically confirmed group and its clinically characterized but neuropathologically unconfirmed clinical group (Table 8). The 1,225 participants in the neuropathologically confirmed cohort included 366 AD dementia cases and 859 cognitively unimpaired and neuropathologically unaffected controls. The 2,610 participants in the clinically classified but neuropathologically unconfirmed cohort included 770 probable AD dementia cases and 1,840 cognitively unimpaired controls.

Ⅳ-3-2. Association of APOE carriers compared to the APOE e3/e3 genotype

Table 9 shows ORs of AD for each APOE genotype and allelic doses after adjustment for age and sex in the neuro-pathologically confirmed and unconfirmed groups, compared to the common APOE e3/e3 genotype. ORs associated with APOE e2 allelic dose and APOE e4 allelic dose were estimated using allelic association tests in an additive genetic model. Since the low frequency of APOE e2/e2, ORs were not showed the expected level of significance in genotypic associations. However, the significance of APOE e2 was observed in the allelic associations among neuro-pathologically confirmed group with OR=0.457. In the case of APOE

e2/e3, the significance was observed in the neuro-pathologically confirmed group than in the suspected group.

The APOE carrier analysis combining the allele types ensured meaningful results and accurate contrasts. The results of the association analysis in the APOE e2 carrier confirmed that the OR was 0.42 times lower than the e3/e3 in the confirmed group. Furthermore, these contrasts were revealed more clearly when the adjusted variables were added, confirming that the risk for the e2 carrier was 0.39 times lower.

IV-3-3. Ages at Alzheimer's disease dementia onset

Age of onset of AD by APOE allele type shows a clear difference. The Kaplan-Meier curve, using a neuro-pathologically confirmed group, shows that the age of onset is higher in groups with e2 than others (Figure 12). Subjects with APOE e2/e2 will present with AD at an estimated mean age of 83 years, while those subject with APOE e3/e3 will present at 81.58 years. This curve also observes the effect of e4 together, subject with APOE e4/e4 will present with AD at an estimated mean age of 72.22 years.

The Kaplan–Meier plots confirmed a relationship between APOE e2 carrier and AD onset ages in older (Figure 13). Compare to the onset age of the subject with APOE e3/e3 (estimated mean ages = 81.58 ± 0.79 in e3/e3), the onset age of group with APOE e2 is delayed by 3 years (estimated mean ages = 84.0 ± 1.7 in e2 carrier). In contrast, the onset age of group with APOE e4 accelerates about 5 years compare to the group with APOE e3/e3 (estimated mean ages = 79.07 ±0.4 in e4 carrier).

IV-3-4. Amyloid accumulation for APOE genotype

The aspect of amyloid accumulation in neuro-pathologically confirmed group was shown to differ depending on APOE. Amyloid accumulation in subjects with APOE e2 showed an unchanged pattern with age, and the estimated regression equation was constructed: $\hat{y} = 1.07 - 4.15 \times 10^{-4}x$. The amyloid accumulation in

subjects with APOE e3/e3 increased slightly with age, and the estimated regression equations was constructed: $\hat{y} = 0.97 + 1.48 \times 10^{-3}x$. The amyloid accumulation in subjects with APOE e4 highly increased with age, and the estimated regression equations was constructed: $\hat{y} = 0.58 + 7.83 \times 10^{-3}x$.

Table 9. Association of APOE genotypes and allelic doses compared to APOE e3/e3.

		Suspected group				Pathologically confirmed group			
	APOE	CN/AD	OR	95%CI	P	CN/AD	OR	95%CI	P
genotypic	e3/e3	1442/448	ref	ref	ref	1775/524	ref	ref	ref
	e2/e2	4/2	0.877	0.09-8.13	0.91	3/1	1.498	0.13-17.03	0.74
	e2/e3	219/57	0.894	0.64-1.23	0.50	86/10	0.351	0.15-0.80	0.01
	e2/e4	23/10	1.423	0.65-3.09	0.37	16/6	1.599	0.53-4.79	0.40
	e3/e4	290/262	3.264	2.64-4.02	2.05E-28	273/165	2.926	2.14-3.99	1.14E-11
	e4/e4	4/49	52.319	18.59-147.17	6.42E-14	6/36	27.939	10.29-75.82	6.27E-11
allelic	e2	223/59	0.897	0.66-1.23	0.50	89/11	0.457	0.22-0.94	0.03
	e4	294/311	3.837	3.18-4.62	2.45E-45	279/201	3.466	2.65-4.54	1.47E-19

For genotypic association tests, odds ratio (OR), 95% confidence interval (CI), and P value (P) for each APOE genotype compared to the APOE e3/e3 genotype were calculated under a logistic regression model.

For allelic association tests, OR, CI, and P associated with APOE e2 allelic dose in APOE e4 non-carriers and APOE e4 allelic dose in APOE e2 non-carriers in an additive genetic model were generated under a logistic regression model.

Table 10. Association of APOE carriers compared to the APOE e3/e3 genotype.

	Suspected group				Confirmed group			
APOE	CN/AD	OR	95%CI	P	CN/AD	OR	95%CI	P
without adj								
e3/e3	1442/448	ref	ref	ref	557/131	ref	ref	ref
e2 carrier	223/59	0.85	0.62-1.17	0.34	223/59	0.42	0.19-0.9	0.03
e4 carrier	294/311	3.65	2.98-4.44	2.21E-37	294/311	3.06	2.29-4.07	1.95E-14
with adj								
e3/e3	1442/448	ref	ref	ref	557/131	ref	ref	ref
e2 carrier	223/59	0.89	0.64-1.23	0.49	223/59	0.39	0.18-0.85	0.02
e4 carrier	294/311	3.88	3.16-4.75	4.03E-39	294/311	3.44	2.54-4.65	9.12E-16

For genotypic association tests, odds ratio (OR), 95% confidence interval (CI), and P value (P) for each APOE genotype compared to the APOE e3/e3 genotype were calculated under a logistic regression model.

Abbreviations: CN, Cognitive normal; AD, Alzheimer's disease dementia; adj, adjustment.

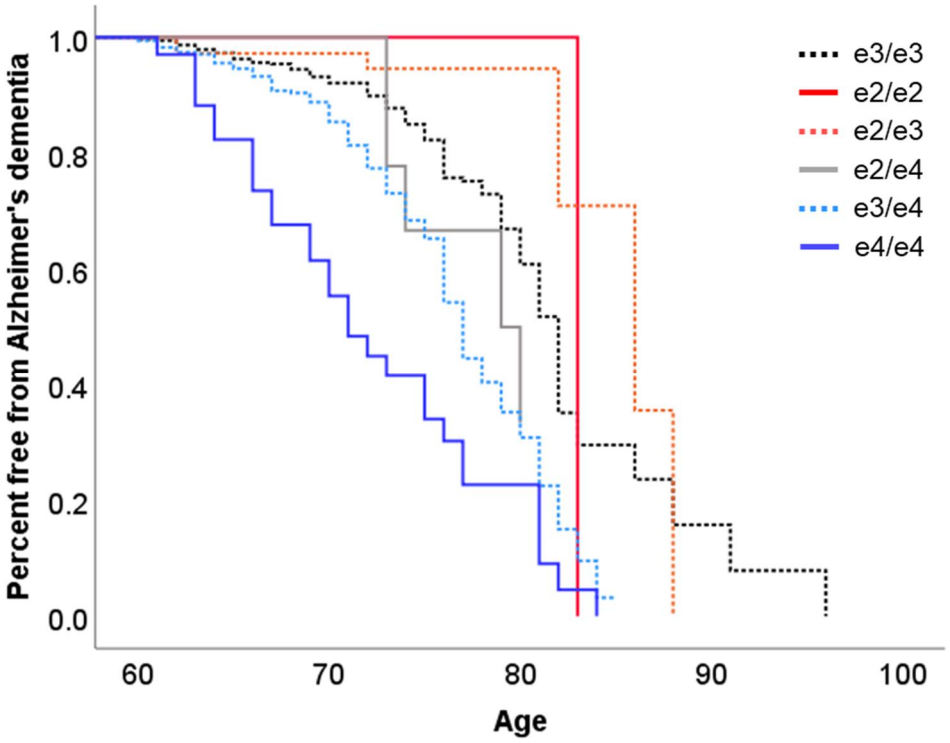


Figure 12. **Percentage of individuals without AD according to age.** Kaplan-Meier curves were generated from AD cases and CN controls in the confirmed group. Y-axis represents the percentage of individuals with each APOE genotype in the pathologically confirmed group. X-axis denotes age ate at onset of cases.

Table 11. Estimated mean of AD onset ages by APOE allele types.

APOE	Mean	SE	95% CI		P
			Lower	Upper	
e2/e2	83.00	0.0	83.0	83.0	
e2/e3	84.73	1.14	82.47	86.95	
e3/e3	81.58	0.79	80.00	83.09	1.62×10^{-20}
e2/e4	77.61	1.14	75.63	79.70	
e3/e4	76.52	0.39	75.76	77.30	
e4/e4	72.22	1.21	69.89	74.61	

P-value was calculated under log rank test.

Abbreviations: SE, standard error; CI, confidence interval.

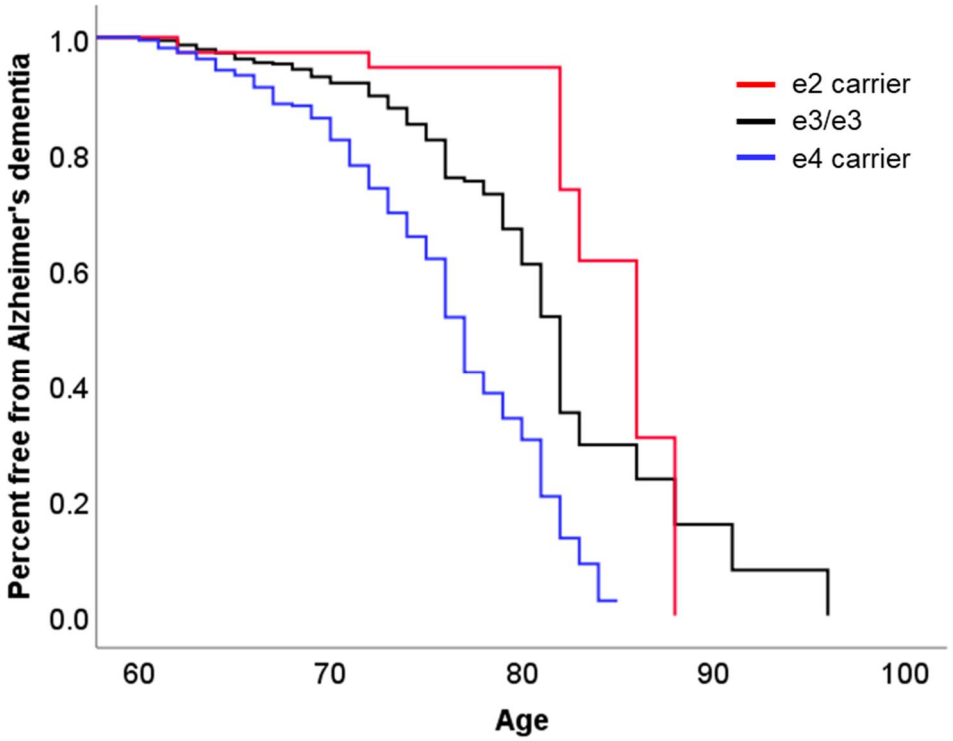


Figure 13. **Percentage of individuals without AD according to age among APOE carriers.** Kaplan-Meier curves were generated from AD cases and CN controls in the confirmed group. Y-axis represents the percentage of individuals with each APOE carrier in the pathologically confirmed group. X-axis denotes age ate at onset of cases.

Table 12. Estimated mean of AD onset ages by APOE carriers.

APOE	Mean	SE	95% CI		P
			Lower	Upper	
e2 carrier	84.41	1.07	82.31	86.52	
e3/e3	81.58	0.79	80.02	83.14	9.01×10^{-19}
e4 carrier	76.07	0.37	75.33	76.80	

P-value was calculated under log rank test.

Abbreviations: SE, standard error; CI, confidence interval.

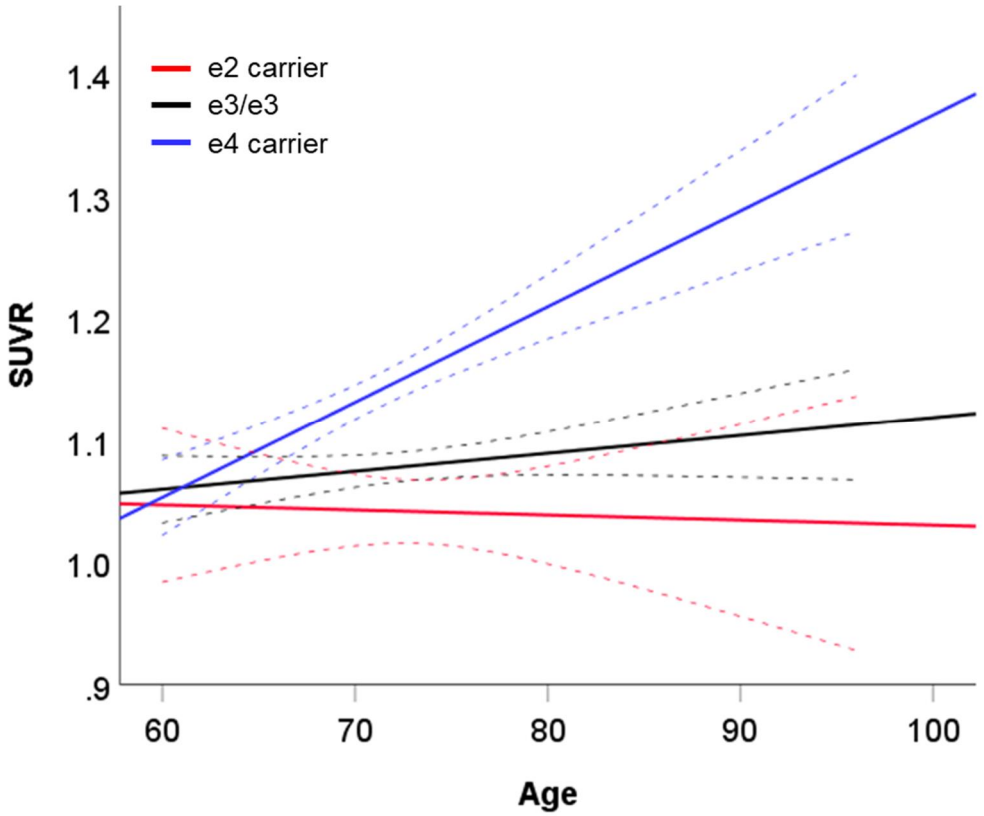


Figure 14. Different changes for standardized uptake value ratio (SUVR) with APOE allele. Y-axis represents the amyloid accumulation score using SUVR from amyloid PET image. X-axis denotes age ate death for controls and age at onset of cases, while replacing with age at death when age at onset was unavailable.

Ⅳ-4. Discussion

This study demonstrates an extremely low probability of AD in APOE e2 in GARD cohort of clinically and neuro-pathologically diagnosed cases and controls. The APOE e2 carrier was associated with a significantly lower 0.39 OR of AD compared to the most common APOE e3/e3 genotype (95% CI: 0.18–0.85) in those neuro-pathologically confirmed subjects. The subject with APOE e2 have 3 years delay in the onset of AD compared to subject with e3/e3. Also, APOE e2 carrier group tend to have a decrease compared to other groups in amyloid accumulation. The findings emphasize the impact of APOE and its variants on AD risk.

Association analysis in the neuro-pathologically confirmed group established a much larger effect of APOE e2 than in the neuro-pathologically unconfirmed group. The low risk and significance of AD in suspected group can probably be attributed to the mix of patients with asymptomatic or multi-causal dementia.

APOE e2/e4, APOE e3/e4, and APOE e4/e4 genotypes compared to APOE e3/e3 genotype in the neuro-pathologically confirmed group were associated with 1.60, 2.93, and 27.94 ORs, respectively. In the case of APOE e2/e4, the protective effect of APOE e2 and risk effect of APOE e4 were conflicting directions, and low toxicity is observed. These results indicate the toxicity of APOE e4 for AD.

Age at disease onset and amyloid accumulation were observed to differ by APOE genotype, and APOE e2 played a protective role against AD. The difference in age at disease onset between the e2 and e4 subjects was almost 8 years. Also the amyloid accumulation pattern in APOE e2 carriers appeared to be unchanged or decreased compared with that in APOE e4 carriers.

The low frequency of the APOE e2/e2 is a limitation in this study to confirm the effect of e2 allele for AD. However, the protective effect of APOE e2 was confirmed indirectly through allelic and carrier strata analysis. By increasing the

number of e2-bearing subjects, the effect of APOE e2 is expected to be further expanded to ensure significance.

Additional research is needed to clarify the mechanism by which APOE and its variants contribute to the pathogenesis and potential treatment and prevention of AD. There is a critical need to discover treatments that account for impact of genotypes on the differential risk of AD dementia, including those that may account for a profound resistance to AD dementia in APOE e2 homozygotes, and to establish their value in the treatment and prevention of AD.

V. 적 요

전장유전체 연관분석을 통한 동아시아인 특이 알츠하이머병 연관 유전변이 발굴 연구

강 사 랑

지도교수 : 이 건 호

조선대학교 대학원

글로벌바이오융합학과

알츠하이머병은 복합적이고 다요인적인 신경퇴행성 질환이며, 노인성 치매의 원인 질병 중 70% 이상을 차지한다. 알츠하이머병은 뇌에서 베타 아밀로이드 단백질의 침착과 신경섬유다발 엉킴과 더불어 신경 세포의 사멸 등에 의해 나이가 진행됨에 따라 인지능력이 감소되는 특징이 있다. 알츠하이머병의 높은 유전률(최대 79%)은 알츠하이머병에 대한 유전학적 연구를 활발하게 하는 근원이 되었으며, 전장유전체 연관분석법은 알츠하이머병의 원인 유전자 발굴을 가속화시켰다. 대규모의 집단 분석 및 메타 분석을 이용한 전장유전체 연관분석을 통해 *APOE* 를 비롯한 40 개 이상의 유전자들이 보고되었다. 그러나, 유전적으로 복잡하고 인종적 발병 차이가 나타나는 알츠하이머병을 이해하기에는 부족하고, 보고된 연구들 대부분이 서양인중심으로 이루어졌다. 본 연구는 동양인 알츠하이머병의 새로운 유전자를 발굴하고 유전적 메커니즘을 규명하는 연구이다.

본 연구는 광주 치매코호트에서 모집된 한국인 자료를 이용하여 전장유전체 연관 분석을 수행하였다. 첫번째로 한국인 2,291 명을 이용한 전장

유전체 연관성 분석 결과, *APOE* 유전자가 가장 강력한 알츠하이머병 위험 유전 변이임을 확인하였다. 뿐만 아니라 서양인 자료에서 기 보고된 치매 연관 유전자 *ABCA7, BIN1* 역시 한국인 자료에서 연관성이 있음을 확인하였다. *APOE* 의 유전적 영향을 효과적으로 배제하기 위한 APOE-stratified 전장유전체 분석을 수행하였으며, 후보 유전자군에 대하여 일본인 자료를 이용하였다. APOE non-carrier 유전체 분석결과, *LRIG1* 의 intron 영역에 있는 rs2280575 와 rs113765515, *CACNA1A* 의 intron 영역에 있는 rs189753894 가 알츠하이머병 위험 유전자로써 가능성을 확인하였다. 유전자 발현량 분석을 통해서 *LRIG1* 유전자와 *CACNA1A* 유전자가 뇌영역에서 높게 발현하고 있음을 확인하였다. 뿐만 아니라, 차등 유전자 발현 분석을 통하여 *LRIG1* 유전자가 entorhinal, frontal, temporal cortex 영역에서 알츠하이머병 진행도에 따라 차등 발현된다는 것을 확인하였다. *CACNA1A* 는 entorhinal cortex 영역에서 차등 발현됨을 확인하였다. 따라서, *LRIG1* 과 *CACNA1A* 유전자는 동아시아인 특이적으로 알츠하이머병과 연관되었음을 시사한다. 두번째로 Bootstrap-based 전장유전체 분석에는 랜덤샘플링을 통한 100 번의 반복 분석이 수행되었으며, 732 개의 유전자 후보군이 추출되었다. 이들 중 아밀로이드 베타 축적과 관련된 유전자를 선별하기 위하여 Drosophila 모델이 이용되었으며, 총 115 개의 유전자가 아밀로이드 축적과 관련이 있음을 확인하였다. pathway 분석을 통하여 42 개의 유전자가 membrane trafficking pathway 에 관련이 있었으며, 그 중 알츠하이머병과의 연관성이 보고되지 않은 새로운 유전자 *AP3D1, CCDC22, DNM3, SH3GL2, PIK3C3* 를 발굴하였다. 이들 유전자는 알츠하이머병 환자의 뇌에서 낮게 발현을 하고 있으며, 베타 아밀로이드 축적과 뇌세포에 영향을 주는 것을 확인하였다. 다음으로 *APOE* e2 에 대한 한국인 알츠하이머병과의 연관성과 위험도를 측정하였다. 연관성 분석을 통해 *APOE* e2 를 한 개 이상 보유하는 집단은 e3/e3 를 보유하고 있는 집단에 비해 알츠하이머병 위험도가

0.39 배 낮다는 것을 확인하였고, 이는 병리학적 진단이 이루어진 그룹에서 더 명확하게 나타난다는 것을 확인하였다. 뿐만 아니라 생존분석을 통하여 e2 를 한 개 이상 보유하는 집단이 e3/e3 에 비하여 알츠하이머 발병 나이가 3 년 늦음을 확인하였다. 나이 진행에 따른 아밀로이드 베타의 축적 정도를 살펴본 결과, e2 를 한 개 이상 보유하는 집단은 그렇지 않은 집단에 비해 아밀로이드 축적에 변화가 없거나 감소하는 것을 확인하였다.

본 연구에서는 대규모 분석을 통해 알츠하이머병과 연관된 새로운 유전자를 추출하였다. 더욱이 본 연구는 동아시아인 특이 단일염기다형성을 보여주었고, 아밀로이드 침착 및 membrane trafficking pathway 관련 유전자를 보여주었으며, APOE 유전자에 대한 알츠하이머병과의 상관성 이해에 중요한 결과이다.

VI. REFERENCES:

ABDULLA, M. A., AHMED, I., ASSAWAMAKIN, A., BHAK, J., BRAHMACHARI, S. K., CALACAL, G. C., CHAURASIA, A., CHEN, C.-H., CHEN, J., CHEN, Y.-T., CHU, J., PAZ, E. M. C. C.-D. L., UNGRIA, M. C. A. D., DELFIN, F. C., EDO, J., FUCHAREON, S., GHANG, H., GOJOBORI, T., HAN, J., HO, S.-F., HOH, B. P., HUANG, W., INOKO, H., JHA, P., JINAM, T. A., JIN, L., JUNG, J., KANGWANPONG, D., KAMPUANSAI, J., KENNEDY, G. C., KHURANA, P., KIM, H.-L., KIM, K., KIM, S., KIM, W.-Y., KIMM, K., KIMURA, R., KOIKE, T., KULAWONGANUNCHAI, S., KUMAR, V., LAI, P. S., LEE, J.-Y., LEE, S., LIU, E. T., MAJUMDER, P. P., MANDAPATI, K. K., MARZUKI, S., MITCHELL, W., MUKERJI, M., NARITOMI, K., NGAMPHIW, C., NIIKAWA, N., NISHIDA, N., OH, B., OH, S., OHASHI, J., OKA, A., ONG, R., PADILLA, C. D., PALITTAPONGARNPIM, P., PERDIGON, H. B., PHIPPS, M. E., PNG, E., SAKAKI, Y., SALVADOR, J. M., SANDRALING, Y., SCARIA, V., SEIELSTAD, M., SIDEK, M. R., SINHA, A., SRIKUMMOOL, M., SUDOYO, H., SUGANO, S., SURYADI, H., SUZUKI, Y., TABBADA, K. A., TAN, A., TOKUNAGA, K., TONGSIMA, S., VILLAMOR, L. P., WANG, E., WANG, Y., WANG, H., WU, J.-Y., XIAO, H., XU, S., YANG, J. O., SHUGART, Y. Y., YOO, H.-S., YUAN, W., ZHAO, G. & ZILFALIL, B. A. 2009. Mapping Human Genetic Diversity in Asia. 326**,** 1541-1545.

AIDARALIEVA, N. J., KAMINO, K., KIMURA, R., YAMAMOTO, M., MORIHARA, T., KAZUI, H., HASHIMOTO, R., TANAKA, T., KUDO, T., KIDA, T., OKUDA, J.-I., UEMA, T., YAMAGATA, H., MIKI, T., AKATSU, H., KOSAKA, K. & TAKEDA, M. 2008. Dynamin 2 gene is a novel susceptibility gene for late-onset Alzheimer disease in non-APOE-ε4 carriers. *Journal of Human Genetics,* 53**,** 296-302.

ALBENSI, B. C. & MATTSON, M. P. 2000. Evidence for the involvement of TNF and NF-κB in hippocampal synaptic plasticity. 35**,** 151-159.

ALSINA, F. C., HITA, F. J., FONTANET, P. A., IRALA, D., HEDMAN, H., LEDDA, F. & PARATCHA, G. 2016. Lrig1 is a cell-intrinsic modulator of hippocampal dendrite complexity and BDNF signaling. *EMBO Rep,* 17**,** 601-16.

ANDERSON, N. B., BULATAO, R. A., COHEN, B., ON RACE, P. & COUNCIL, N. R. 2004. Ethnic differences in dementia and Alzheimer's disease. *Critical perspectives on racial and ethnic differences in health in late life.* National Academies Press (US).

ARBET, J., MCGUE, M., CHATTERJEE, S. & BASU, S. 2017. Resampling-based tests for Lasso in genome-wide association studies. *BMC Genetics,* 18**,** 70.

AUTON, A., BRYC, K., BOYKO, A. R., LOHMUELLER, K. E., NOVEMBRE, J., REYNOLDS, A., INDAP, A., WRIGHT, M. H., DEGENHARDT, J. D., GUTENKUNST, R. N., KING, K. S., NELSON, M. R. & BUSTAMANTE, C. D. 2009. Global distribution of genomic diversity underscores rich complex history of continental human populations. *Genome Res,* 19**,** 795-803.

BIER, E. 2005. Drosophila, the golden bug, emerges as a tool for human genetics. *Nature Reviews Genetics,* 6**,** 9-23.

BLATTLER, A., YAO, L., WITT, H., GUO, Y., NICOLET, C. M., BERMAN, B. P. & FARNHAM, P. J. 2014. Global loss of DNA methylation uncovers intronic enhancers in genes showing expression changes. *Genome Biol,* 15**,** 469.

BLOOM, G. S. 2014. Amyloid-β and tau: the trigger and bullet in Alzheimer disease pathogenesis. *JAMA Neurol,* 71**,** 505-8.

BRAMHAM, C. R. & MESSAOUDI, E. 2005. BDNF function in adult synaptic plasticity: the synaptic consolidation hypothesis. *Prog Neurobiol,* 76**,** 99-125.

BUCHMAN, A. S., YU, L., BOYLE, P. A., SCHNEIDER, J. A., DE JAGER, P. L. & BENNETT, D. A. 2016. Higher brain BDNF gene expression is associated with slower cognitive decline in older adults. *Neurology,* 86**,** 735-41.

CHANG, C. C., CHOW, C. C., TELLIER, L. C., VATTIKUTI, S., PURCELL, S. M. & LEE, J. J. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience,* 4.

CONSORTIUM, G., ARDLIE, K. G., DELUCA, D. S., SEGRê, A. V., SULLIVAN, T. J., YOUNG, T. R., GELFAND, E. T., TROWBRIDGE, C. A., MALLER, J. B. & TUKIAINEN, T. J. S. 2015. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. 348**,** 648-660.

CORDER, E. H., SAUNDERS, A. M., RISCH, N. J., STRITTMATTER, W. J., SCHMECHEL, D. E., GASKELL, P. C., JR., RIMMLER, J. B., LOCKE, P. A., CONNEALLY, P. M., SCHMADER, K. E. & ET AL. 1994. Protective effect of apolipoprotein E type 2 allele for late onset Alzheimer disease. *Nat Genet,* 7**,** 180-4.

DALE, A. M., FISCHL, B. & SERENO, M. I. 1999. Cortical Surface-Based Analysis: I. Segmentation and Surface Reconstruction. *NeuroImage,* 9**,** 179-194.

DE LEEUW, C. A., MOOIJ, J. M., HESKES, T. & POSTHUMA, D. J. P. C. B. 2015. MAGMA: generalized gene-set analysis of GWAS data. 11**,** e1004219.

DEMURO, A., SMITH, M. & PARKER, I. 2011. Single-channel Ca(2+) imaging implicates Aβ1-42 amyloid pores in Alzheimer's disease pathology. *The Journal of cell biology,* 195**,** 515-524.

EPIS, R., MARCELLO, E., GARDONI, F. & DI LUCA, M. 2012. Alpha, beta-and gamma-secretases in Alzheimer's disease. *Front Biosci (Schol Ed),* 4**,** 1126-50.

FARRER, L. A., CUPPLES, L. A., HAINES, J. L., HYMAN, B., KUKULL, W. A., MAYEUX, R., MYERS, R. H., PERICAK-VANCE, M. A., RISCH, N. & VAN DUIJN, C. M. 1997. Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis. APOE and Alzheimer Disease Meta Analysis Consortium. *Jama,* 278**,** 1349-56.

FAYE, L. L., SUN, L., DIMITROMANOLAKIS, A. & BULL, S. B. 2011. A flexible genome-wide bootstrap method that accounts for ranking and threshold-selection bias in GWAS interpretation and replication study design. *Stat Med,* 30**,** 1898-912.

FISCHL, B. & DALE, A. M. 2000. Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proceedings of the National Academy of Sciences of the United States of America,* 97**,** 11050-11055.

FISCHL, B., SERENO, M. I. & DALE, A. M. 1999. Cortical Surface-Based Analysis: II: Inflation, Flattening, and a Surface-Based Coordinate System. *NeuroImage,* 9**,** 195-207.

FREEDMAN, M. L., MONTEIRO, A. N., GAYTHER, S. A., COETZEE, G. A., RISCH, A., PLASS, C., CASEY, G., DE BIASI, M., CARLSON, C., DUGGAN, D., JAMES, M., LIU, P., TICHELAAR, J. W., VIKIS, H. G., YOU, M. & MILLS, I. G. 2011. Principles for the post-GWAS functional characterization of cancer risk loci. *Nat Genet,* 43**,** 513-8.

GATZ, M., REYNOLDS, C. A., FRATIGLIONI, L., JOHANSSON, B., MORTIMER, J. A., BERG, S., FISKE, A. & PEDERSEN, N. L. 2006. Role of Genes and Environments for Explaining Alzheimer Disease. *Archives of General Psychiatry,* 63**,** 168-174.

GHANI, M., REITZ, C., CHENG, R., VARDARAJAN, B. N., JUN, G., SATO, C., NAJ, A., RAJBHANDARY, R., WANG, L.-S., VALLADARES, O., LIN, C.-F., LARSON, E. B., GRAFF-RADFORD, N. R., EVANS, D., DE JAGER, P. L., CRANE, P. K., BUXBAUM, J. D., MURRELL, J. R., RAJ, T., ERTEKIN-TANER, N., LOGUE, M., BALDWIN, C. T., GREEN, R. C., BARNES, L. L., CANTWELL, L. B., FALLIN, M. D., GO, R. C. P., GRIFFITH, P. A., OBISESAN, T. O., MANLY, J. J., LUNETTA, K. L., KAMBOH, M. I., LOPEZ, O. L., BENNETT, D. A., HENDRIE, H., HALL, K. S., GOATE, A. M., BYRD, G. S., KUKULL, W. A., FOROUD, T. M., HAINES, J. L., FARRER, L. A., PERICAK-VANCE, M. A., LEE, J. H., SCHELLENBERG, G. D., ST. GEORGE-HYSLOP, P., MAYEUX, R., ROGAEVA, E. & CONSORTIUM, F. T. A. S. D. G. 2015. Association of Long Runs of Homozygosity With Alzheimer Disease Among African American Individuals. *JAMA Neurology,* 72**,** 1313-1323.

GUARDIA, C. M., DE PACE, R., MATTERA, R. & BONIFACINO, J. S. 2018. Neuronal functions of adaptor complexes involved in protein sorting. *Current Opinion in Neurobiology,* 51**,** 103-110.

GURKOFF, G., SHAHLAIE, K., LYETH, B. & BERMAN, R. 2013. Voltage-gated calcium channel antagonists and traumatic brain injury. *Pharmaceuticals (Basel),* 6**,** 788-812.

HAROLD, D., ABRAHAM, R., HOLLINGWORTH, P., SIMS, R., GERRISH, A., HAMSHERE, M. L., PAHWA, J. S., MOSKVINA, V., DOWZELL, K., WILLIAMS, A., JONES, N., THOMAS, C., STRETTON, A., MORGAN, A. R., LOVESTONE, S., POWELL, J., PROITSI, P., LUPTON, M. K., BRAYNE, C., RUBINSZTEIN, D. C., GILL, M., LAWLOR, B., LYNCH, A., MORGAN, K., BROWN, K. S., PASSMORE, P. A., CRAIG, D., MCGUINNESS, B., TODD, S., HOLMES, C., MANN, D., SMITH, A. D., LOVE, S., KEHOE, P. G., HARDY, J., MEAD, S., FOX, N., ROSSOR, M., COLLINGE, J., MAIER, W., JESSEN, F., SCHþRMANN, B., HEUN, R., VAN DEN BUSSCHE, H., HEUSER, I., KORNHUBER, J., WILTFANG, J., DICHGANS, M., FRøLICH, L., HAMPEL, H., HþLL, M., RUJESCU, D., GOATE, A. M., KAUWE, J. S. K., CRUCHAGA, C., NOWOTNY, P., MORRIS, J. C., MAYO, K., SLEEGERS, K., BETTENS, K., ENGELBORGHS, S., DE DEYN, P. P., VAN BROECKHOVEN, C., LIVINGSTON, G., BASS, N. J., GURLING, H., MCQUILLIN, A., GWILLIAM, R., DELOUKAS, P., AL-CHALABI, A., SHAW, C. E., TSOLAKI, M., SINGLETON, A. B., GUERREIRO, R., MþHLEISEN, T. W., NøTHEN, M. M., MOEBUS, S., JøCKEL, K.-H., KLOPP, N., WICHMANN, H. E., CARRASQUILLO, M. M., PANKRATZ, V. S., YOUNKIN, S. G., HOLMANS, P. A., O'DONOVAN, M., OWEN, M. J. & WILLIAMS, J. 2009. Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nature Genetics,* 41**,** 1088-1093.

HOLLINGWORTH, P., HAROLD, D., SIMS, R., GERRISH, A., LAMBERT, J.-C., CARRASQUILLO, M. M., ABRAHAM, R., HAMSHERE, M. L., PAHWA, J. S., MOSKVINA, V., DOWZELL, K., JONES, N., STRETTON, A., THOMAS, C., RICHARDS, A., IVANOV, D., WIDDOWSON, C., CHAPMAN, J., LOVESTONE, S., POWELL, J., PROITSI, P., LUPTON, M. K., BRAYNE, C., RUBINSZTEIN, D. C., GILL, M., LAWLOR, B., LYNCH, A., BROWN, K. S., PASSMORE, P. A., CRAIG, D., MCGUINNESS, B., TODD, S., HOLMES, C., MANN, D., SMITH, A. D., BEAUMONT, H., WARDEN, D., WILCOCK, G., LOVE, S., KEHOE, P. G., HOOPER, N. M., VARDY, E. R. L. C., HARDY, J., MEAD, S., FOX, N. C., ROSSOR, M., COLLINGE, J., MAIER, W., JESSEN, F., RþTHER, E., SCHþRMANN, B., HEUN, R., KøLSCH, H., VAN DEN BUSSCHE, H., HEUSER, I., KORNHUBER, J., WILTFANG, J., DICHGANS, M., FRøLICH, L., HAMPEL, H., GALLACHER, J., HþLL, M., RUJESCU, D., GIEGLING, I., GOATE, A. M., KAUWE, J. S. K., CRUCHAGA, C., NOWOTNY, P., MORRIS, J. C., MAYO, K., SLEEGERS, K., BETTENS, K., ENGELBORGHS, S., DE DEYN, P. P., VAN BROECKHOVEN, C., LIVINGSTON, G., BASS, N. J., GURLING, H., MCQUILLIN, A., GWILLIAM, R., DELOUKAS, P., AL-CHALABI, A., SHAW, C. E., TSOLAKI, M., SINGLETON, A. B., GUERREIRO, R., MþHLEISEN, T. W., NøTHEN, M. M., MOEBUS, S., JøCKEL, K.-H., KLOPP, N., WICHMANN, H. E., PANKRATZ, V. S., SANDO, S. B., AASLY, J. O., BARCIKOWSKA, M., WSZOLEK, Z. K., DICKSON, D. W., GRAFF-RADFORD, N. R., PETERSEN, R. C., et al. 2011. Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. *Nature Genetics,* 43**,** 429-435.

HOLLINGWORTH, P., SWEET, R., SIMS, R., HAROLD, D., RUSSO, G., ABRAHAM, R., STRETTON, A., JONES, N., GERRISH, A., CHAPMAN, J., IVANOV, D., MOSKVINA, V., LOVESTONE, S., PRIOTSI, P., LUPTON, M., BRAYNE, C., GILL, M., LAWLOR, B., LYNCH, A., CRAIG, D., MCGUINNESS, B., JOHNSTON, J., HOLMES, C., LIVINGSTON, G., BASS, N. J., GURLING, H., MCQUILLIN, A., HOLMANS, P., JONES, L., DEVLIN, B., KLEI, L., BARMADA, M. M., DEMIRCI, F. Y., DEKOSKY, S. T., LOPEZ, O. L., PASSMORE, P., OWEN, M. J., O'DONOVAN, M. C., MAYEUX, R., KAMBOH, M. I. & WILLIAMS, J. 2012. Genome-wide association study of Alzheimer's disease with psychotic symptoms. *Mol Psychiatry,* 17**,** 1316-27.

HOWELL, G. J., HOLLOWAY, Z. G., COBBOLD, C., MONACO, A. P. & PONNAMBALAM, S. 2006. Cell biology of membrane trafficking in human disease. *International review of cytology,* 252**,** 1-69.

ITKIN, A., DUPRES, V., DUFRìNE, Y. F., BECHINGER, B., RUYSSCHAERT, J.-M. & RAUSSENS, V. 2011. Calcium ions promote formation of amyloid β-peptide (1-40) oligomers causally implicated in neuronal toxicity of Alzheimer's disease. *PloS one,* 6**,** e18250-e18250.

JANSEN, I. E., SAVAGE, J. E., WATANABE, K., BRYOIS, J., WILLIAMS, D. M., STEINBERG, S., SEALOCK, J., KARLSSON, I. K., HæGG, S., ATHANASIU, L., VOYLE, N., PROITSI, P., WITOELAR, A., STRINGER, S., AARSLAND, D., ALMDAHL, I. S., ANDERSEN, F., BERGH, S., BETTELLA, F., BJORNSSON, S., BRèKHUS, A., BRçTHEN, G., DE LEEUW, C., DESIKAN, R. S., DJUROVIC, S., DUMITRESCU, L., FLADBY, T., HOHMAN, T. J., JONSSON, P. V., KIDDLE, S. J., RONGVE, A., SALTVEDT, I., SANDO, S. B., SELBèK, G.,

SHOAI, M., SKENE, N. G., SNAEDAL, J., STORDAL, E., ULSTEIN, I. D., WANG, Y., WHITE, L. R., HARDY, J., HJERLING-LEFFLER, J., SULLIVAN, P. F., VAN DER FLIER, W. M., DOBSON, R., DAVIS, L. K., STEFANSSON, H., STEFANSSON, K., PEDERSEN, N. L., RIPKE, S., ANDREASSEN, O. A. & POSTHUMA, D. 2019. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nature Genetics,* 51**,** 404-413.

JANSEN, W. J., OSSENKOPPELE, R., KNOL, D. L., TIJMS, B. M., SCHELTENS, P., VERHEY, F. R. J., VISSER, P. J. & GROUP, A. T. A. B. S. 2015. Prevalence of Cerebral Amyloid Pathology in Persons Without Dementia: A Meta-analysis. *JAMA,* 313**,** 1924-1938.

JIAO, S. S., SHEN, L. L., ZHU, C., BU, X. L., LIU, Y. H., LIU, C. H., YAO, X. Q., ZHANG, L. L., ZHOU, H. D., WALKER, D. G., TAN, J., GøTZ, J., ZHOU, X. F. & WANG, Y. J. 2016. Brain-derived neurotrophic factor protects against tau-related neurodegeneration of Alzheimer's disease. *Transl Psychiatry,* 6**,** e907.

JIN, C., ZHANG, L., XIAN, Y., LIU, X., WU, Y., ZHANG, F., ZHU, J., ZHANG, G., CHEN, C., GONG, R., ZHANG, L., YUAN, J., ZHANG, F., TIAN, L., WANG, G. & CHENG, Z. 2014. The SORL1 polymorphism rs985421 may confer the risk for amnestic mild cognitive impairment and Alzheimer's disease in the Han Chinese population. *Neurosci Lett,* 563**,** 80-4.

JOHNSON, J. L. & ABECASIS, G. R. J. B. 2017. GAS Power Calculator: web-based power calculator for genetic association studies. 164343.

JUN, G. R., CHUNG, J., MEZ, J., BARBER, R., BEECHAM, G. W., BENNETT, D. A., BUXBAUM, J. D., BYRD, G. S., CARRASQUILLO, M. M., CRANE, P. K., CRUCHAGA, C., DE JAGER, P., ERTEKIN-TANER, N., EVANS, D., FALLIN, M. D., FOROUD, T. M., FRIEDLAND, R. P., GOATE, A. M., GRAFF-RADFORD, N. R., HENDRIE, H., HALL, K. S., HAMILTON-NELSON, K. L., INZELBERG, R., KAMBOH, M. I., KAUWE, J. S. K., KUKULL, W. A., KUNKLE, B. W., KUWANO, R., LARSON, E. B., LOGUE, M. W., MANLY, J. J., MARTIN, E. R., MONTINE, T. J., MUKHERJEE, S., NAJ, A., REIMAN, E. M., REITZ, C., SHERVA, R., ST GEORGE-HYSLOP, P. H., THORNTON, T., YOUNKIN, S. G., VARDARAJAN, B. N., WANG, L. S., WENDLUND, J. R., WINSLOW, A. R., HAINES, J., MAYEUX, R., PERICAK-VANCE, M. A., SCHELLENBERG, G., LUNETTA, K. L. & FARRER, L. A. 2017. Transethnic genome-wide scan identifies novel Alzheimer's disease loci. *Alzheimers Dement,* 13**,** 727-738.

KANEKIYO, T., XU, H. & BU, G. 2014. ApoE and Aβ in Alzheimer's disease: accidental encounters or partners? *Neuron,* 81**,** 740-754.

KANG, Y., NA, D., HAHN, S. J. I. H. B. R. & CO, C. 2003. Seoul neuropsychological screening battery.

KANG, Y., NA, D. L. & HAHN, S. J. J. O. T. K. N. A. 1997. A validity study on the Korean Mini-Mental State Examination (K-MMSE) in dementia patients. 15**,** 300-308.

KIM, Y. J., SEO, S. W., PARK, S. B., YANG, J. J., LEE, J. S., LEE, J., JANG, Y. K., KIM, S. T., LEE, K. H., LEE, J. M., LEE, J. H., KIM, J. S., NA, D. L. & KIM, H. J. 2017. Protective effects of APOE e2 against disease progression in subcortical vascular mild cognitive impairment patients: A three-year longitudinal study. *Sci Rep,* 7**,** 1910.

KUNKLE, B. W., GRENIER-BOLEY, B., SIMS, R., BIS, J. C., DAMOTTE, V., NAJ, A. C., BOLAND, A., VRONSKAYA, M., VAN DER LEE, S. J., AMLIE-WOLF, A., BELLENGUEZ, C., FRIZATTI, A., CHOURAKI, V., MARTIN, E. R., SLEEGERS, K., BADARINARAYAN, N., JAKOBSDOTTIR, J., HAMILTON-NELSON, K. L., MORENO-GRAU, S., OLASO, R., RAYBOULD, R., CHEN, Y., KUZMA, A. B., HILTUNEN, M., MORGAN, T., AHMAD, S., VARDARAJAN, B. N., EPELBAUM, J., HOFFMANN, P., BOADA, M., BEECHAM, G. W., GARNIER, J. G., HAROLD, D., FITZPATRICK, A. L., VALLADARES, O., MOUTET, M. L., GERRISH, A., SMITH, A. V., QU, L., BACQ, D., DENNING, N., JIAN, X., ZHAO, Y., DEL ZOMPO, M., FOX, N. C., CHOI, S. H., MATEO, I., HUGHES, J. T., ADAMS, H. H., MALAMON, J., SANCHEZ-GARCIA, F., PATEL, Y., BRODY, J. A., DOMBROSKI, B. A., NARANJO, M. C. D., DANIILIDOU, M., EIRIKSDOTTIR, G., MUKHERJEE, S., WALLON, D., UPHILL, J., ASPELUND, T., CANTWELL, L. B., GARZIA, F., GALIMBERTI, D., HOFER, E., BUTKIEWICZ, M., FIN, B., SCARPINI, E., SARNOWSKI, C., BUSH, W. S., MESLAGE, S., KORNHUBER, J., WHITE, C. C., SONG, Y., BARBER, R. C., ENGELBORGHS, S., SORDON, S., VOIJNOVIC, D., ADAMS, P. M., VANDENBERGHE, R., MAYHAUS, M., CUPPLES, L. A., ALBERT, M. S., DE DEYN, P. P., GU, W., HIMALI, J. J., BEEKLY, D., SQUASSINA, A., HARTMANN, A. M., ORELLANA, A., BLACKER, D., RODRIGUEZ-RODRIGUEZ, E., LOVESTONE, S., GARCIA, M. E., DOODY, R. S., MUNOZ-FERNADEZ, C., SUSSAMS, R., LIN, H., FAIRCHILD, T. J., BENITO, Y. A., et al. 2019. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. *Nat Genet,* 51**,** 414-430.

KUO, C.-L., PILLING, L. C., ATKINS, J. L., KUCHEL, G. A. & MELZER, D. 2020. ApoE e2 and aging-related outcomes in 379,000 UK Biobank participants. 2020.02.12.20022459.

LAMBERT, J. C., IBRAHIM-VERBAAS, C. A., HAROLD, D., NAJ, A. C., SIMS, R., BELLENGUEZ, C., DESTAFANO, A. L., BIS, J. C., BEECHAM, G. W., GRENIER-BOLEY, B., RUSSO, G., THORTON-WELLS, T. A., JONES, N., SMITH, A. V., CHOURAKI, V., THOMAS, C., IKRAM, M. A., ZELENIKA, D., VARDARAJAN, B. N., KAMATANI, Y., LIN, C. F., GERRISH, A., SCHMIDT, H., KUNKLE, B., DUNSTAN, M. L., RUIZ, A., BIHOREAU, M. T., CHOI, S. H., REITZ, C., PASQUIER, F., CRUCHAGA, C., CRAIG, D., AMIN, N., BERR, C., LOPEZ, O. L., DE JAGER, P. L., DERAMECOURT, V., JOHNSTON, J. A., EVANS, D., LOVESTONE, S., LETENNEUR, L., MORõN, F. J., RUBINSZTEIN, D. C., EIRIKSDOTTIR, G., SLEEGERS, K., GOATE, A. M., FIëVET, N., HUENTELMAN, M. W., GILL, M., BROWN, K., KAMBOH, M. I., KELLER, L., BARBERGER-GATEAU, P., MCGUINESS, B., LARSON, E. B., GREEN, R., MYERS, A. J., DUFOUIL, C., TODD, S., WALLON, D., LOVE, S., ROGAEVA, E., GALLACHER, J., ST GEORGE-HYSLOP, P., CLARIMON, J., LLEO, A., BAYER, A., TSUANG, D. W., YU, L., TSOLAKI, M., BOSSû, P., SPALLETTA, G., PROITSI, P., COLLINGE, J., SORBI, S., SANCHEZ-GARCIA, F., FOX, N. C., HARDY, J., DENIZ NARANJO, M. C., BOSCO, P., CLARKE, R., BRAYNE, C., GALIMBERTI, D., MANCUSO, M., MATTHEWS, F., MOEBUS, S., MECOCCI, P., DEL ZOMPO, M., MAIER, W., HAMPEL, H., PILOTTO, A., BULLIDO, M., PANZA, F., CAFFARRA, P., NACMIAS, B., GILBERT, J. R., MAYHAUS, M., LANNEFELT, L., HAKONARSON, H., PICHLER, S., et al.

2013. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet,* 45**,** 1452-8.

LASKE, C., STELLOS, K., HOFFMANN, N., STRANSKY, E., STRATEN, G., ESCHWEILER, G. W. & LEYHE, T. 2011. Higher BDNF serum levels predict slower cognitive decline in Alzheimer's disease patients. *Int J Neuropsychopharmacol,* 14**,** 399-404.

LEE, J., KANG, M., LEE, O., LEE, H., KWAK, O. & YOO, W. 2021. Korean Dementia Observatory 2020. Seoul: National Institute of Dementia; 2021. Report No. NIDR-2002-0031.

LIZIO, M., HARSHBARGER, J., SHIMOJI, H., SEVERIN, J., KASUKAWA, T., SAHIN, S., ABUGESSAISA, I., FUKUDA, S., HORI, F. & ISHIKAWA-KATO, S. J. G. B. 2015. Gateways to the FANTOM5 promoter level mammalian expression atlas. 16**,** 1-14.

LU, B., NAGAPPAN, G., GUAN, X., NATHAN, P. J. & WREN, P. 2013. BDNF-based synaptic repair as a disease-modifying strategy for neurodegenerative diseases. *Nat Rev Neurosci,* 14**,** 401-16.

MAASS, A., LANDAU, S., BAKER, S. L., HORNG, A., LOCKHART, S. N., LA JOIE, R., RABINOVICI, G. D., JAGUST, W. J. & ALZHEIMER'S DISEASE NEUROIMAGING, I. 2017. Comparison of multiple tau-PET measures as biomarkers in aging and Alzheimer's disease. *NeuroImage,* 157**,** 448-463.

MARTINS, I. J., HONE, E., FOSTER, J. K., SþNRAM-LEA, S. I., GNJEC, A., FULLER, S. J., NOLAN, D., GANDY, S. E. & MARTINS, R. N. 2006. Apolipoprotein E, cholesterol metabolism, diabetes, and the convergence of risk factors for Alzheimer's disease and cardiovascular disease. *Molecular Psychiatry,* 11**,** 721-736.

MCCARTHY, S., DAS, S., KRETZSCHMAR, W., DELANEAU, O., WOOD, A. R., TEUMER, A., KANG, H. M., FUCHSBERGER, C., DANECEK, P., SHARP, K., LUO, Y., SIDORE, C., KWONG, A., TIMPSON, N., KOSKINEN, S., VRIEZE, S., SCOTT, L. J., ZHANG, H., MAHAJAN, A., VELDINK, J., PETERS, U., PATO, C., VAN DUIJN, C. M., GILLIES, C. E., GANDIN, I., MEZZAVILLA, M., GILLY, A., COCCA, M., TRAGLIA, M., ANGIUS, A., BARRETT, J. C., BOOMSMA, D., BRANHAM, K., BREEN, G., BRUMMETT, C. M., BUSONERO, F., CAMPBELL, H., CHAN, A., CHEN, S., CHEW, E., COLLINS, F. S., CORBIN, L. J., SMITH, G. D., DEDOUSSIS, G., DORR, M., FARMAKI, A.-E., FERRUCCI, L., FORER, L., FRASER, R. M., GABRIEL, S., LEVY, S., GROOP, L., HARRISON, T., HATTERSLEY, A., HOLMEN, O. L., HVEEM, K., KRETZLER, M., LEE, J. C., MCGUE, M., MEITINGER, T., MELZER, D., MIN, J. L., MOHLKE, K. L., VINCENT, J. B., NAUCK, M., NICKERSON, D., PALOTIE, A., PATO, M., PIRASTU, N., MCINNIS, M., RICHARDS, J. B., SALA, C., SALOMAA, V., SCHLESSINGER, D., SCHOENHERR, S., SLAGBOOM, P. E., SMALL, K., SPECTOR, T., STAMBOLIAN, D., TUKE, M., TUOMILEHTO, J., VAN DEN BERG, L. H., VAN RHEENEN, W., VOLKER, U., WIJMENGA, C., TONIOLO, D., ZEGGINI, E., GASPARINI, P., SAMPSON, M. G., WILSON, J. F., FRAYLING, T., DE BAKKER, P. I. W., SWERTZ, M. A., MCCARROLL, S., KOOPERBERG, C., DEKKER, A., ALTSHULER, D., WILLER, C., IACONO, W., RIPATTI, S., et al. 2016. A reference panel of 64,976 haplotypes for genotype imputation. *Nature Genetics,* 48**,** 1279-1283.

MCKHANN, G., DRACHMAN, D., FOLSTEIN, M., KATZMAN, R., PRICE, D. & STADLAN, E. M. 1984. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology,* 34**,** 939-44.

MENG, Y., BALDWIN, C. T., BOWIRRAT, A., WARASKA, K., INZELBERG, R., FRIEDLAND, R. P. & FARRER, L. A. 2006. Association of polymorphisms in the Angiotensin-converting enzyme gene with Alzheimer disease in an Israeli Arab community. *Am J Hum Genet,* 78**,** 871-877.

MIRANDA, A. M. & DI PAOLO, G. 2018. Endolysosomal dysfunction and exosome secretion: implications for neurodegenerative disorders. *Cell Stress,* 2**,** 115-118.

MIYASHITA, A., KOIKE, A., JUN, G., WANG, L.-S., TAKAHASHI, S., MATSUBARA, E., KAWARABAYASHI, T., SHOJI, M., TOMITA, N., ARAI, H., ASADA, T., HARIGAYA, Y., IKEDA, M., AMARI, M., HANYU, H., HIGUCHI, S., IKEUCHI, T., NISHIZAWA, M., SUGA, M., KAWASE, Y., AKATSU, H., KOSAKA, K., YAMAMOTO, T., IMAGAWA, M., HAMAGUCHI, T., YAMADA, M., MORIAHA, T., TAKEDA, M., TAKAO, T., NAKATA, K., FUJISAWA, Y., SASAKI, K., WATANABE, K., NAKASHIMA, K., URAKAMI, K., OOYA, T., TAKAHASHI, M., YUZURIHA, T., SERIKAWA, K., YOSHIMOTO, S., NAKAGAWA, R., KIM, J.-W., KI, C.-S., WON, H.-H., NA, D. L., SEO, S. W., MOOK-JUNG, I., THE ALZHEIMER DISEASE GENETICS, C., ST. GEORGE-HYSLOP, P., MAYEUX, R., HAINES, J. L., PERICAK-VANCE, M. A., YOSHIDA, M., NISHIDA, N., TOKUNAGA, K., YAMAMOTO, K., TSUJI, S., KANAZAWA, I., IHARA, Y., SCHELLENBERG, G. D., FARRER, L. A. & KUWANO, R. 2013. SORL1 Is Genetically Associated with Late-Onset Alzheimer's Disease in Japanese, Koreans and Caucasians. *PLOS ONE,* 8**,** e58618.

MOON, S., KIM, Y. J., HAN, S., HWANG, M. Y., SHIN, D. M., PARK, M. Y., LU, Y., YOON, K., JANG, H.-M., KIM, Y. K., PARK, T.-J., SONG, D. S., PARK, J. K., LEE, J.-E. & KIM, B.-J. 2019. The Korea Biobank Array: Design and Identification of Coding Variants Associated with Blood Biochemical Traits. *Scientific Reports,* 9**,** 1382.

MOREL, E., CHAMOUN, Z., LASIECKA, Z. M., CHAN, R. B., WILLIAMSON, R. L., VETANOVETZ, C., DALL'ARMI, C., SIMOES, S., POINT DU JOUR, K. S., MCCABE, B. D., SMALL, S. A. & DI PAOLO, G. 2013. Phosphatidylinositol-3-phosphate regulates sorting and processing of amyloid precursor protein through the endosomal system. *Nature Communications,* 4**,** 2250.

MUKHERJEE, S., KIM, S., RAMANAN, V. K., GIBBONS, L. E., NHO, K., GLYMOUR, M. M., ERTEKIN-TANER, N., MONTINE, T. J., SAYKIN, A. J., CRANE, P. K. & FOR THE ALZHEIMER'S DISEASE NEUROIMAGING, I. 2014. Gene-based GWAS and biological pathway analysis of the resilience of executive functioning. *Brain Imaging and Behavior,* 8**,** 110-118.

NAM, H., HWANG, S., KIM, Y., BYON, S. & KIM, K. 2017. Korean dementia observatory 2017. Seongnam: Central Dementia Center; 2017. Report No. NIDR-1704-0019.

NG, T. K. S., HO, C. S. H., TAM, W. W. S., KUA, E. H. & HO, R. C. 2019. Decreased Serum Brain-Derived Neurotrophic Factor (BDNF) Levels in Patients with Alzheimer's Disease (AD): A Systematic Review and Meta-Analysis. *Int J Mol Sci,* 20.

OLIVER, D. & REDDY, P. H. 2019. Dynamics of Dynamin-Related Protein 1 in Alzheimer's Disease and Other Neurodegenerative Diseases. *Cells,* 8.

PALMQVIST, S., SCHøLL, M., STRANDBERG, O., MATTSSON, N., STOMRUD, E., ZETTERBERG, H., BLENNOW, K., LANDAU, S., JAGUST, W. & HANSSON, O. 2017. Earliest accumulation of β-amyloid occurs within the default-mode network and concurrently affects brain connectivity. *Nature Communications,* 8**,** 1214.

PATEL, H., HODGES, A. K., CURTIS, C., LEE, S. H., TROAKES, C., DOBSON, R. J. B. & NEWHOUSE, S. J. 2019. Transcriptomic analysis of probable asymptomatic and symptomatic alzheimer brains. *Brain, Behavior, and Immunity,* 80**,** 644-656.

PATTERSON, N., PRICE, A. L. & REICH, D. J. P. G. 2006. Population structure and eigenanalysis. 2**,** e190.

PHAN, L., JIN, Y., ZHANG, H., QIANG, W., SHEKHTMAN, E., SHAO, D., REVOE, D., VILLAMARIN, R., IVANCHENKO, E. & KIMURA, M. J. N. C. F. B. I., US NATIONAL LIBRARY OF MEDICINE 2020. ALFA: Allele Frequency Aggregator. 10.

PINI, L., PIEVANI, M., BOCCHETTA, M., ALTOMARE, D., BOSCO, P., CAVEDO, E., GALLUZZI, S., MARIZZONI, M. & FRISONI, G. B. 2016. Brain atrophy in Alzheimer's Disease and aging. *Ageing Res Rev,* 30**,** 25-48.

PRICE, A. L., PATTERSON, N. J., PLENGE, R. M., WEINBLATT, M. E., SHADICK, N. A. & REICH, D. 2006. Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics,* 38**,** 904-909.

RABER, J., HUANG, Y. & ASHFORD, J. W. 2004. ApoE genotype accounts for the vast majority of AD risk and AD pathology. *Neurobiol Aging,* 25**,** 641-50.

RAJAN, K. B., BARNES, L. L., WILSON, R. S., MCANINCH, E. A., WEUVE, J., SIGHOKO, D. & EVANS, D. A. 2017. Racial Differences in the Association Between Apolipoprotein E Risk Alleles and Overall and Total Cardiovascular Mortality Over 18 Years. 65**,** 2425-2430.

RAJAN, K. B., WEUVE, J., BARNES, L. L., MCANINCH, E. A., WILSON, R. S., EVANS, D. A. J. A. S. & DEMENTIA 2021. Population estimate of people with clinical Alzheimer's disease and mild cognitive impairment in the United States (2020–2060).

RAJENDRAN, L. & ANNAERT, W. 2012. Membrane trafficking pathways in Alzheimer's disease. *Traffic,* 13**,** 759-70.

REIMAN, E. M., ARBOLEDA-VELASQUEZ, J. F., QUIROZ, Y. T., HUENTELMAN, M. J., BEACH, T. G., CASELLI, R. J., CHEN, Y., SU, Y., MYERS, A. J., HARDY, J., PAUL VONSATTEL, J., YOUNKIN, S. G., BENNETT, D. A., DE JAGER, P. L., LARSON, E. B., CRANE, P. K., KEENE, C. D., KAMBOH, M. I., KOFLER, J. K., DUQUE, L., GILBERT, J. R., GWIRTSMAN, H. E., BUXBAUM, J. D., DICKSON, D. W., FROSCH, M. P., GHETTI, B. F., LUNETTA, K. L., WANG, L. S., HYMAN, B. T., KUKULL, W. A., FOROUD, T., HAINES, J. L., MAYEUX, R. P., PERICAK-VANCE, M. A., SCHNEIDER, J. A., TROJANOWSKI, J. Q., FARRER, L. A., SCHELLENBERG, G. D., BEECHAM, G. W., MONTINE, T. J. & JUN, G. R. 2020. Exceptionally low likelihood of Alzheimer's dementia in APOE2 homozygotes from a 5,000-person neuropathological study. *Nat Commun,* 11**,** 667.

REN, Y., XU, H. W., DAVEY, F., TAYLOR, M., AITON, J., COOTE, P., FANG, F., YAO, J., CHEN, D., CHEN, J. X., YAN, S. D. & GUNN-MOORE, F. J. 2008. Endophilin I Expression Is Increased in the Brains of Alzheimer Disease Patients *. *Journal of Biological Chemistry,* 283**,** 5685-5691.

RIDGE, P. G., MUKHERJEE, S., CRANE, P. K., KAUWE, J. S. K. & ALZHEIMER'S DISEASE GENETICS, C. 2013. Alzheimer's Disease: Analyzing the Missing Heritability. *PLOS ONE,* 8**,** e79771.

SHASTRY, B. S. 2002. SNP alleles in human disease and evolution. *Journal of Human Genetics,* 47**,** 561-566.

SHEN, L. & JIA, J. 2016. An Overview of Genome-Wide Association Studies in Alzheimer's Disease. *Neurosci Bull,* 32**,** 183-90.

SHERVA, R., BALDWIN, C. T., INZELBERG, R., VARDARAJAN, B., CUPPLES, L. A., LUNETTA, K., BOWIRRAT, A., NAJ, A., PERICAK-VANCE, M., FRIEDLAND, R. P. & FARRER, L. A. 2011. Identification of novel candidate genes for Alzheimer's disease by autozygosity mapping using genome wide SNP data. *J Alzheimers Dis,* 23**,** 349-59.

SHINOHARA, M., KANEKIYO, T., TACHIBANA, M., KURTI, A., SHINOHARA, M., FU, Y., ZHAO, J., HAN, X., SULLIVAN, P. M., REBECK, G. W., FRYER, J. D., HECKMAN, M. G. & BU, G. 2020. APOE2 is associated with longevity independent of Alzheimer's disease. *Elife,* 9.

SHULMAN, J. M., CHIPENDO, P., CHIBNIK, L. B., AUBIN, C., TRAN, D., KEENAN, B. T., KRAMER, P. L., SCHNEIDER, J. A., BENNETT, D. A., FEANY, M. B. & DE JAGER, P. L. 2011. Functional screening of Alzheimer pathology genome-wide association signals in Drosophila. *American journal of human genetics,* 88**,** 232-238.

SINGH, P. P., SINGH, M. & MASTANA, S. S. 2006. APOE distribution in world populations with new data from India and the UK. *Annals of Human Biology,* 33**,** 279-308.

STAROKADOMSKYY, P., GLUCK, N., LI, H., CHEN, B., WALLIS, M., MAINE, G. N., MAO, X., ZAIDI, I. W., HEIN, M. Y., MCDONALD, F. J., LENZNER, S., ZECHA, A., ROPERS, H. H., KUSS, A. W., MCGAUGHRAN, J., GECZ, J. & BURSTEIN, E. 2013. CCDC22 deficiency in humans blunts activation of proinflammatory NF-κB signaling. *J Clin Invest,* 123**,** 2244-56.

STRITTMATTER, W. J., SAUNDERS, A. M., SCHMECHEL, D., PERICAK-VANCE, M., ENGHILD, J., SALVESEN, G. S. & ROSES, A. D. 1993. Apolipoprotein E: high-avidity binding to beta-amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease. *Proc Natl Acad Sci U S A,* 90**,** 1977-81.

SUN, Y., ZHU, J., ZHOU, D., CANCHI, S., WU, C., COX, N. J., RISSMAN, R. A., GAMAZON, E. R. & WU, L. 2021. A transcriptome-wide association study of Alzheimer's disease using prediction models of relevant tissues identifies novel candidate susceptibility genes. *Genome Medicine,* 13**,** 141.

SWEIGART, B., ANDERSEN, S. L., GURINOVICH, A., COSENTINO, S., SCHUPF, N., PERLS, T. T. & SEBASTIANI, P. 2021. APOE E2/E2 Is Associated with Slower Rate of Cognitive Decline with Age. *J Alzheimers Dis,* 83**,** 853-860.

THIES, W. & BLEILER, L. J. A. D. 2021. 2021 Alzheimer's disease facts and figures. 17**,** 327-406.

TOSTO, G., FU, H., VARDARAJAN, B. N., LEE, J. H., CHENG, R., REYES-DUMEYER, D., LANTIGUA, R., MEDRANO, M., JIMENEZ-VELAZQUEZ, I. Z., ELKIND, M. S., WRIGHT, C. B., SACCO, R. L., PERICAK-VANCE, M., FARRER, L., ROGAEVA, E., ST GEORGE-HYSLOP, P., REITZ, C. & MAYEUX, R. 2015. F-box/LRR-repeat protein 7 is genetically associated with Alzheimer's disease. *Ann Clin Transl Neurol,* 2**,** 810-20.

UEMURA, K., KUZUYA, A. & SHIMOHAMA, S. 2004. Protein trafficking and Alzheimer's disease. *Curr Alzheimer Res,* 1**,** 1-10.

UHLëN, M., BJøRLING, E., AGATON, C., SZIGYARTO, C. A.-K., AMINI, B., ANDERSEN, E., ANDERSSON, A.-C., ANGELIDOU, P., ASPLUND, A., ASPLUND, C. J. M. & PROTEOMICS, C. 2005. A human protein atlas for normal and cancer tissues based on antibody proteomics. 4**,** 1920-1932.

VëLEZ, J. I., CHANDRASEKHARAPPA, S. C., HENAO, E., MARTINEZ, A. F., HARPER, U., JONES, M., SOLOMON, B. D., LOPEZ, L., GARCIA, G., AGUIRRE-ACEVEDO, D. C., ACOSTA-BAENA, N., CORREA, J. C., LOPERA-GõMEZ, C. M., JARAMILLO-ELORZA, M. C., RIVERA, D., KOSIK, K. S., SCHORK, N. J., SWANSON, J. M., LOPERA, F. & ARCOS-BURGOS, M. 2013. Pooling/bootstrap-based GWAS (pbGWAS) identifies new loci modifying the age of onset in PSEN1 p.Glu280Ala Alzheimer's disease. *Molecular psychiatry,* 18**,** 568-575.

VëLEZ, J. I., LOPERA, F., SEPULVEDA-FALLA, D., PATEL, H. R., JOHAR, A. S., CHUAH, A., TOBõN, C., RIVERA, D., VILLEGAS, A., CAI, Y., PENG, K., ARKELL, R., CASTELLANOS, F. X., ANDREWS, S. J., SILVA LARA, M. F., CREAGH, P. K., EASTEAL, S., DE LEON, J., WONG, M. L., LICINIO, J., MASTRONARDI, C. A. & ARCOS-BURGOS, M. 2016. APOE*E2 allele delays age of onset in PSEN1 E280A Alzheimer's disease. *Mol Psychiatry,* 21**,** 916-24.

WANG, D., CHAN, C.-C., CHERRY, S. & HIESINGER, P. R. 2013. Membrane trafficking in neuronal maintenance and degeneration. *Cellular and Molecular Life Sciences,* 70**,** 2919-2934.

WANG, L., BUDOLFSON, K. & WANG, F. 2011. Pik3c3 deletion in pyramidal neurons results in loss of synapses, extensive gliosis and progressive neurodegeneration. *Neuroscience,* 172**,** 427-442.

WANGLER, M. F., HU, Y. & SHULMAN, J. M. 2017. Drosophila and genome-wide association studies: a review and resource for the functional dissection of human complex traits. *Dis Model Mech,* 10**,** 77-88.

WIGHTMAN, D. P., JANSEN, I. E., SAVAGE, J. E., SHADRIN, A. A., BAHRAMI, S., HOLLAND, D., RONGVE, A., BúRTE, S., WINSVOLD, B. S., DRANGE, O. K., MARTINSEN, A. E., SKOGHOLT, A. H., WILLER, C., BRçTHEN, G., BOSNES, I., NIELSEN, J. B., FRITSCHE, L. G., THOMAS, L. F., PEDERSEN, L. M., GABRIELSEN, M. E., JOHNSEN, M. B., MEISINGSET, T. W., ZHOU, W., PROITSI, P., HODGES, A., DOBSON, R., VELAYUDHAN, L., HEILBRON, K., AUTON, A., SEALOCK, J. M., DAVIS, L. K., PEDERSEN, N. L., REYNOLDS, C. A., KARLSSON, I. K., MAGNUSSON, S., STEFANSSON, H., THORDARDOTTIR, S., JONSSON, P. V., SNAEDAL, J., ZETTERGREN, A., SKOOG, I., KERN, S., WAERN, M., ZETTERBERG, H., BLENNOW, K., STORDAL, E., HVEEM, K., ZWART, J. A., ATHANASIU, L., SELNES, P., SALTVEDT, I.,

SANDO, S. B., ULSTEIN, I., DJUROVIC, S., FLADBY, T., AARSLAND, D., SELBèK, G., RIPKE, S., STEFANSSON, K., ANDREASSEN, O. A. & POSTHUMA, D. 2021. A genome-wide association study with 1,126,563 individuals identifies new risk loci for Alzheimer's disease. *Nat Genet,* 53**,** 1276-1282.

WILLER, C. J., LI, Y. & ABECASIS, G. R. J. B. 2010. METAL: fast and efficient meta-analysis of genomewide association scans. 26**,** 2190-2191.

WINKLE, C. C. & GUPTON, S. L. 2016. Membrane Trafficking in Neuronal Development: Ins and Outs of Neural Connectivity. *International review of cell and molecular biology,* 322**,** 247-280.

XIE, A.-J., HOU, T.-Y., XIONG, W., HUANG, H.-Z., ZHENG, J., LI, K., MAN, H.-Y., HU, Y.-Z., HAN, Z.-T., ZHANG, H.-H., WEI, N., WANG, J.-Z., LIU, D., LU, Y. & ZHU, L.-Q. 2019. Tau overexpression impairs neuronal endocytosis by decreasing the GTPase dynamin 1 through the miR-132/MeCP2 pathway. *Aging cell,* 18**,** e12929-e12929.

XU, F., VITEK, M. P., COLTON, C. A., PREVITI, M. L., DAVIS, J. & VAN NOSTRAND, W. E. 2012. Human apolipoprotein E2 promotes parenchymal amyloid deposition and neuronal loss in vasculotropic mutant amyloid-β protein Tg-SwDI mice. *J Alzheimers Dis,* 31**,** 359-69.

YANG, J., LEE, S. H., GODDARD, M. E. & VISSCHER, P. M. J. T. A. J. O. H. G. 2011. GCTA: a tool for genome-wide complex trait analysis. 88**,** 76-82.

YANG, Y., CHEN, J., GUO, Z., DENG, S., DU, X., ZHU, S., YE, C., SHI, Y. S. & LIU, J.-J. 2018. Endophilin A1 Promotes Actin Polymerization in Dendritic Spines Required for Synaptic Potentiation. 11.

YANG, Y., WEI, M., XIONG, Y., DU, X., ZHU, S., YANG, L., ZHANG, C. & LIU, J.-J. 2015. Endophilin A1 regulates dendritic spine morphogenesis and stability through interaction with p140Cap. *Cell Research,* 25**,** 496-516.

YU, Q., WANG, Y., DU, F., YAN, S., HU, G., ORIGLIA, N., RUTIGLIANO, G., SUN, Q., YU, H., AINGE, J., YAN, S. F., GUNN-MOORE, F. & YAN, S. S. 2018. Overexpression of endophilin A1 exacerbates synaptic alterations in a mouse model of Alzheimer's disease. *Nature Communications,* 9**,** 2968.

ZHANG, C.-C., WANG, H.-F., TAN, M.-S., WAN, Y., ZHANG, W., ZHENG, Z.-J., KONG, L.-L., WANG, Z.-X., TAN, L. & JIANG, T. J. M. N. 2017. SORL1 is associated with the risk of late-onset Alzheimer's disease: a replication study and meta-analyses. 54**,** 1725-1732.

ZHOU, T., THUNG, K. H., LIU, M. & SHEN, D. 2019. Brain-Wide Genome-Wide Association Study for Alzheimer's Disease via Joint Projection Learning and Sparse Regression Model. *IEEE Transactions on Biomedical Engineering,* 66**,** 165-175.

ZHOU, X., CHEN, Y., MOK, K. Y., ZHAO, Q., CHEN, K., CHEN, Y., HARDY, J., LI, Y., FU, A. K. Y., GUO, Q., IP, N. Y. & INITIATIVE, F. T. A. S. D. N. 2018. Identification of genetic risk factors in the Chinese population implicates a role of immune system in Alzheimer's disease pathogenesis. 115**,** 1697-1706.

ZHU, L., SU, M., LUCAST, L., LIU, L., NETZER, W. J., GANDY, S. E. & CAI, D. 2012. Dynamin 1 Regulates Amyloid Generation through Modulation of BACE-1. *PLOS ONE,* 7**,** e45033.

VII. APPENDIX

VII-1. Scripts for analysis

VII-1-1. Analysis.sh

```
#you need pheno.txt and smpl.txt per each analysis
#make symdata file
ln -s
/tera/home/sarang/Data/Chosun5545_dnalinkQC_rmregion_HRC1.1imputed/chosu
n_5545_dnalinkqc_rmregion_HRC-1.1_imputed_chr1-
22_re4/chosun_5545_dnalinkqc_rmregion_HRC-1.1_imputed_chr1-
22_re4_qcmaf.bed plink.bed
ln -s
/tera/home/sarang/Data/Chosun5545_dnalinkQC_rmregion_HRC1.1imputed/chosu
n_5545_dnalinkqc_rmregion_HRC-1.1_imputed_chr1-
22_re4/chosun_5545_dnalinkqc_rmregion_HRC-1.1_imputed_chr1-
22_re4_qcmaf.bim plink.bim
ln -s
/tera/home/sarang/Data/Chosun5545_dnalinkQC_rmregion_HRC1.1imputed/chosu
n_5545_dnalinkqc_rmregion_HRC-1.1_imputed_chr1-
22_re4/chosun_5545_dnalinkqc_rmregion_HRC-1.1_imputed_chr1-
22_re4_qcmaf.fam plink.fam
mkdir Heritability
mkdir ./Logistic/jap
#logistic regression
plink --bfile plink --hide-covar --logistic --maf 0.01 --pheno ./info/pheno.txt --
keep ./info/smpl.txt --out ./Logistic/logistic --covar ./info/cov_sexage4pc.txt --ci
0.95
```

```
#draw manhattan & qq plot
#Rscript
/tera/home/sarang/Data/Chosun5545_dnalinkQC_rmregion_HRC1.1imputed/qqma
nhat-v2.R ./Logistic/logistic.assoc.logistic
#extract p<0.0001
awk 'NR==1 {print}; $12 < 0.0001' ./Logistic/logistic.assoc.logistic
> ./Logistic/top_logi_00001
#make snplist p<0.0001
awk '{print $2}' ./Logistic/top_logi_00001 > ./Logistic/top_snp_00001
#annotate snplist
grep -F -w -f ./Logistic/top_snp_00001
/tera/home/sarang/chosun_3719_160928_maf0.01_ibd_filt_0.9_HRC_imputed_chr
1-22_vcf_annot_info-1.csv > ./Logistic/annot_top_snp_00001
#Minor allele freq
plink --bfile plink --freq case-control --out ./Logistic/freq --pheno ./info/pheno.txt -
-extract ./Logistic/top_snp_00001
#match with jap result
grep -F -w -f ./Logistic/top_snp_00001 ./Logistic/jap/logistic_ci95.assoc.logistic
> ./Logistic/top_snp_jap
#meta_p00001
plink --meta-
analysis ./Logistic/top_logi_00001 ./Logistic/jap/logistic_ci95.assoc.logistic --
out ./Logistic/meta_p00001
#summary results
Rscript result_sum2.R
```

VII-1-2. qqmanhat-v2.R

```
#!/usr/bin/env Rscript
```

```
args = commandArgs(trailingOnly=TRUE)
library("snpStats")
library("qqman")
gwas <- read.table(args[1], header=TRUE)
ylimit <- as.numeric(args[2])
## calculate lambda
qq <- qq.chisq(-2 * log(gwas$P), df=2, xlab="Expected", ylab="Observed",
conc=c(0.05,0.95), thin=c(0.25,50), overdisp=TRUE, pvals=TRUE, pch=20,
col.shade="gray", trim=0.5)
mat <- as.matrix(print(qq))
lam <- format(round((mat[3,1]), 4), nsmall=2)
lam1 <- paste("Lambda =",lam)
## plot qq
outpdfqq <- paste(args[1],"_qqplot.pdf", sep="")
outpngqq <- paste(args[1],"_qqplot.png", sep="")
outpdfman <- paste(args[1],"_manplot.pdf", sep="")
outpngman <- paste(args[1],"_manplot.png", sep="")
pdf(outpdfqq, width=12, height=9)
qq.chisq(-2 * log(gwas$P), df=2, xlab="Expected", ylab="Observed",
conc=c(0.05,0.95), thin=c(0.25,50), overdisp=TRUE, pvals=TRUE, pch=20,
col.shade="gray", trim=0.5)
legend(20,14, lam1, cex=0.8, bty="n")
png(filename = outpngqq, width=840, height=733, units = "px", res=92)
qq.chisq(-2 * log(gwas$P), df=2, xlab="Expected", ylab="Observed",
conc=c(0.05,0.95), thin=c(0.25,50), overdisp=TRUE, pvals=TRUE, pch=20,
col.shade="gray", trim=0.5)
legend(20,14, lam1, cex=0.8, bty="n")
```

```
## plot manhattan
#pdf(outpdfman, width=12, height=9)
#manhattan(gwas, ylim=c(0,10), chr = "CHR", bp = "BP", p = "P", snp = "SNP",
col = c("orange",
"darkblue","firebrick3","forestgreen","maroon4","deepskyblue","gray28","hotpink"
,"olivedrab3","black","orchid4"), chrlabs = NULL, suggestiveline = -log10(1e-05),
highlight = NULL, logp = TRUE, annotatePval = 0.00001, annotateTop = TRUE)
png(filename = outpngman, width=1280, height=680, units = "px", res=92)
manhattan(gwas, ylim=c(0,16), chr = "CHR", bp = "BP", p = "P", snp = "SNP", col
= c("orange",
"darkblue","firebrick3","forestgreen","maroon4","deepskyblue","gray28","hotpink"
,"olivedrab3","black","orchid4"), chrlabs = NULL, suggestiveline = -log10(1e-05),
genomewideline = -log10(5e-08), highlight = NULL, logp = TRUE, annotatePval =
0.00001, annotateTop = TRUE)
#manhattan(gwas, ylim=c(0,ylimit), chr = "CHR", bp = "BP", p = "P", snp =
"SNP", col = c("orange",
"darkblue","firebrick3","forestgreen","maroon4","deepskyblue","gray28","hotpink"
,"olivedrab3","black","orchid4"), chrlabs = NULL, suggestiveline = -log10(5e-06),
genomewideline = -log10(5e-08), highlight = NULL, logp = TRUE)
dev.off()
dev.off()
dev.off()
dev.off()
```

VII-1-3. result_sum2.R

```
# file1 : top_logi_00001          CHR SNP
# file2 : top_snp_jap                   SNP P OR SE L95 U95
# file3 : meta_p00001.meta        SNP P OR Q I
```

```
# file4 : annot_top_00001                    V1 V2 'V3' V4 V5 V6 V7 'V8' V9
library(plyr)
library(tidyr)
library(dplyr)
result <- read.table("./Logistic/top_logi_00001", header=TRUE)
jap <- read.table("./Logistic/top_snp_jap", header=TRUE)
meta <- read.table("./Logistic/meta_p00001.meta", header=TRUE)
freq <- read.table("./Logistic/freq.frq.cc", header=TRUE)
annot <- read.table("./Logistic/annot_top_snp_00001", header=FALSE, sep="\t",
fill=TRUE)
freq <- transform(freq, MAF= ((MAF_A*NCHROBS_A) +
(MAF_U*NCHROBS_U))/(NCHROBS_A+NCHROBS_U))
result$A2 = freq[match(result$SNP, freq$SNP), "A2"]
result$MAF = freq[match(result$SNP, freq$SNP), "MAF"]
result$JAP = jap[(match(result$SNP, jap$SNP)),]
result$META = meta[(match(result$SNP, meta$SNP)),]
result$ANNT = annot[match(result$SNP, annot$V1),]
write.table(result, file="./Logistic/result_sum_top00001_r2", na="NA",
quote=FALSE, row.names=F)
```

VII-1-4. meta-analysis.sh

```
#!/bin/bash
##you need metal_process.txt , convert_OR2beta.R
##sh meta-analysis.sh logistic logistic_ci95
mkdir METAL
cp metal_process.txt ./METAL/
awk -F " " '{print $2}' ./Logistic/"$1".assoc.logistic > ./METAL/tmp_1_snp
awk -F " " '{print $2}' ./Logistic/jap/"$2".assoc.logistic > ./METAL/tmp_2_snp
```

```
awk '{X=""; for (i=1;i<=NF;i+=1) {printf "%s%s", X, $i; X="\t"}; printf
"\n"}' ./Logistic/"$1".assoc.logistic > ./METAL/"$1".assoc.logistic_ADD
awk '{X=""; for (i=1;i<=NF;i+=1) {printf "%s%s", X, $i; X="\t"}; printf
"\n"}' ./Logistic/jap/"$2".assoc.logistic > ./METAL/"$2".assoc.logistic_ADD
awk -F " " '{print
$2"\t"$1"\t"$3"\t"$4"\t"$5"\t"$6"\t"$7"\t"$8"\t"$9"\t"$10"\t"$12}' ./METAL/"$1".a
ssoc.logistic_ADD > ./METAL/tmp_ADD_1
awk -F " " '{print
$2"\t"$1"\t"$3"\t"$4"\t"$5"\t"$6"\t"$7"\t"$8"\t"$9"\t"$10"\t"$12}' ./METAL/"$2".a
ssoc.logistic_ADD > ./METAL/tmp_ADD_2
awk -F " " 'FNR==NR{x2[$1]=$0; next} $1 in x2 {print
x2[$1]}' ./METAL/tmp_ADD_2 ./METAL/tmp_1_snp
> ./METAL/"$2".assoc.logistic_ADD_overlap
awk -F " " '{print $1}' ./METAL/"$2".assoc.logistic_ADD_overlap
> ./METAL/tmp_3_snp
awk -F " " 'FNR==NR{x2[$1]=$0; next} $1 in x2 {print
x2[$1]}' ./METAL/tmp_ADD_1 ./METAL/tmp_3_snp
> ./METAL/"$1".assoc.logistic_ADD_overlap
rm -rf ./METAL/tmp*
# create allele frequency file
plink --bfile plink --freq --out ./METAL/"$1"_freq
plink --bfile
/tera/home/sarang/Data/Chosun5545_dnalinkQC_rmregion_HRC1.1imputed/nigatt
a_maf001_imputed_gwas/nigatta_afterQC_exsnp --freq --out ./METAL/"$2"_freq
# create input files for metal
awk -F " " '{print $2"\t"$3"\t"$4}' ./METAL/"$1"_freq.frq |
        sed 's/A1\tA2/RISK\tREF/g' > ./METAL/tmp_1_freq
awk -F " " '{print $2"\t"$3"\t"$4}' ./METAL/"$2"_freq.frq |
```

```
        sed 's/A1\tA2/RISK\tREF/g' > ./METAL/tmp_2_freq
awk -F " " '{print $1}' ./METAL/"$1".assoc.logistic_ADD_overlap
> ./METAL/tmp_1_snp
awk -F " " '{print $1}' ./METAL/"$2".assoc.logistic_ADD_overlap
> ./METAL/tmp_2_snp
awk -F " " 'FNR==NR{x2[$1]=$0; next} $1 in x2 {print
x2[$1]}' ./METAL/tmp_1_freq ./METAL/tmp_1_snp > ./METAL/tmp_1_freq_1
awk -F " " 'FNR==NR{x2[$1]=$0; next} $1 in x2 {print
x2[$1]}' ./METAL/tmp_2_freq ./METAL/tmp_2_snp > ./METAL/tmp_2_freq_2
paste -d "\t" ./METAL/"$1".assoc.logistic_ADD_overlap ./METAL/tmp_1_freq_1
> ./METAL/"$1".assoc.logistic_ADD_overlap_freq
paste -d "\t" ./METAL/"$2".assoc.logistic_ADD_overlap ./METAL/tmp_2_freq_2
> ./METAL/"$2".assoc.logistic_ADD_overlap_freq
Rscript convert_OR2beta.R ./METAL/"$1".assoc.logistic_ADD_overlap_freq  ##
convert OR to BETA
Rscript convert_OR2beta.R ./METAL/"$2".assoc.logistic_ADD_overlap_freq   ##
convert OR to BETA
sed 's/"//g' ./METAL/"$1".assoc.logistic_ADD_overlap_freq_BETA.txt
> ./METAL/tmp_1_freq_beta_1
sed 's/"//g' ./METAL/"$2".assoc.logistic_ADD_overlap_freq_BETA.txt
> ./METAL/tmp_2_freq_beta_2
mv ./METAL/tmp_1_freq_beta_1 ./METAL/file_1_metal.txt
mv ./METAL/tmp_2_freq_beta_2 ./METAL/file_2_metal.txt &&
rm -rf ./METAL/tmp*
cd METAL
echo "run metal"
####type 'metal' in the terminal####
#SOURCE metal_process.txt
```

VII-1-5. locuszoom.sh

```
#make gene list > genelist.txt
#make SNP"\t"P list > snpP
awk '{print $2"\t"$12}' ./Logistic/logistic.assoc.logistic > ./Logistic/snpP

#for gene
for i in $(cat ./LD/genelist.txt)
do
        mkdir ./LD/$i
        cp ./Assoc/snpP ./LD/$i/snpP
        cd ./LD/$i/
        locuszoom --flank 400kb --metal snpP --markercol SNP --pvalcol P --build
hg19 --pop ASN --source 1000G_Nov2010 --refgene $i --delim space
        cd ..
        cd ..
done

#for single snp with rs#
#mkdir LD
#make locus_top_snp file in LD directory
awk '{print $2" "$9}' ./Assoc/assoc.assoc > ./Assoc/snpP
for i in $(cat ./LD/locus_top_snp)
        do
        mkdir ./LD/$i
        cp ./Assoc/snpP ./LD/$i/snpP
        plink --bfile plink --ld-snp $i --ld-window 99999 --ld-window-kb 400 --ld-
window-r2 0 --r2 dprime with-freqs --out ./LD/$i/ld$i
```

```
        awk 'BEGIN {print "snp1 snp2 dprime rsquare"} {if(NR>1) print $3" "$7"
"$10" "$9}'  ./LD/$i/ld$i.ld  > ./LD/$i/snp12dr2
        cd LD/$i
        locuszoom --metal snpP --markercol SNP --pvalcol P --ld snp12dr2 --refsnp
$i --source 1000G_Nov2010 --build hg19 --pop ASN --delim space
        cd ..
        cd ..
done

#for chr & position
mkdir ./LD/1:1234
plink --bfile plink --ld-snp 1:1234 --ld-window 99999 --ld-window-kb 400 --ld-
window-r2 0 --r2 dprime with-freqs --out ./LD/11234/ld11234
awk 'BEGIN {print "snp1 snp2 dprime rsquare"} {if(NR>1) print $3" "$7" "$10"
"$9}' ./LD/7156258568/ld7156258568.ld  > snp12dr2
locuszoom --metal snpP --markercol SNP --pvalcol P --chr 1 --start 1234-400000 --
end 1234+400000 --source 1000G_Nov2010 --build hg19 --pop ASN --delim space
--ld snp12dr2 --flank 400kb
```

VII-1-5. Analysis_assoc.sh

```
mkdir Assoc
mkdir ./Assoc/jap
mkdir info

ln -s -f /tera/home/sarang/Data/plink_genotype_merged6602/ANA2/plink.bim
plink.bim
ln -s -f /tera/home/sarang/Data/plink_genotype_merged6602/ANA2/plink.fam
plink.fam
```

```
ln -s -f /tera/home/sarang/Data/plink_genotype_merged6602/ANA2/plink.bed
plink.bed

#association test
plink --bfile plink --assoc --out ./Assoc/assoc --pheno ./info/pheno.txt --maf 0.01 --
ci 0.95 --keep ./info/smpl.txt

#extract p<0.0001
awk 'NR==1 {print}; $9 < 0.0001' ./Assoc/assoc.assoc > ./Assoc/top_assoc_00001
#make snplist p<0.0001
awk '{print $2}' ./Assoc/top_assoc_00001 > ./Assoc/top_snp_00001
#annotate (geno)
grep -F -w -f ./Assoc/top_snp_00001
/tera/home/sarang/ANNOTATION/Axiom_Custom_Array.na34.annot.csv
> ./Assoc/annot_top_00001
#annotate (impu)
grep -F -w -f ./Assoc/top_snp_00001
/tera/home/sarang/chosun_3719_160928_maf0.01_ibd_filt_0.9_HRC_imputed_chr
1-22_vcf_annot_info-1.csv > ./Assoc/annot_top_00001

#match with jap result
grep -F -w -f ./Assoc/top_snp_00001 ./Assoc/jap/assoc.assoc
> ./Assoc/top_snp_jap

#meta_p00001
plink --meta-analysis ./Assoc/top_assoc_00001 ./Assoc/jap/assoc.assoc --
out ./Assoc/meta_p00001
```

```
#paste files
Rscript result_sum.R

#draw manhattan plot
Rscript
/tera/home/sarang/Data/Chosun5545_dnalinkQC_rmregion_HRC1.1imputed/qqma
nhat_1.R ./Assoc/assoc.assoc
#draw q-qplot
Rscript
/tera/home/sarang/Data/Chosun5545_dnalinkQC_rmregion_HRC1.1imputed/qqpli
nk.R ./Assoc/assoc.assoc

#Minor allele freq
plink --bfile plink --freq case-control --out ./Assoc/freq --pheno ./info/pheno.txt --
extract ./Assoc/top_snp_00001 --keep ./info/smpl.txt

#Model test for freq (mm/mM/MM)
plink --bfile plink --model --out ./Assoc/model --pheno ./info/pheno.txt --
extract ./Assoc/top_snp_00001 --keep ./info/smpl.txt
awk 'NR==1 {print} $5~/GENO/' ./Assoc/model.model > ./Assoc/model_geno

#ln -s /tera/home/sarang/Data/plink_genotype_merged6602/result_sum.R
result_sum.R
ln -s
/tera/home/sarang/Data/Chosun5545_dnalinkQC_rmregion_HRC1.1imputed/result
_sum.R result_sum_assoc.R

#summary results
```

Rscript result_sum_assoc.R

VII-1-6. result_sum_assoc.R

```
# file1 : top_assoc_00001          CHR SNP
# file2 : top_snp_jap                    SNP P OR SE L95 U95
# file3 : meta_p00001.meta         SNP P OR Q I
# file4 : annot_top_00001                V1 V2 'V3' V4 V5 V6 V7 'V8' V9
library(plyr)
library(tidyr)
library(dplyr)
result <- read.table("./Assoc/top_assoc_00001", header=TRUE)
jap <- read.table("./Assoc/top_snp_jap", header=TRUE)
meta <- read.table("./Assoc/meta_p00001.meta", header=TRUE)
model <- read.table("./Assoc/model_geno", header=T)
annot <- read.table("./Assoc/annot_top_00001", header=FALSE, sep="\t",
fill=TRUE, col.name=c("SNP","ANNT"))
result$FRQ = model[match(result$SNP, model$SNP), c(6,7)]
result$ANNT = annot[match(result$SNP, annot$SNP), "ANNT"]
result2 <- separate(result, ANNT, c(NA,"REGION", NA, "GENE", NA, NA), sep=
"\\|")
result2$JAP = jap[(match(result$SNP, jap$SNP)),]
result2$META = meta[(match(result$SNP, meta$SNP)),]
write.table(result2, file="./Assoc/result_sum_top00001_r", na="NA",
quote=FALSE)
```

VIII. ACKNOWLEDGEMENT

I sincerely thank my supervisor Prof. Kun Ho Lee for his support and guidance in the completion of my doctoral studies. I would also like to thank the rest of my thesis committee members, Prof. Seok-Jun Kim, Prof. Jung-woong Kim, Prof. Han Yong Lee, and Prof. Jungsoo Gim for all the valuable comments and suggestions.

My special thanks for unwavering support form the people in Gwangju Alzheimer's & Related Dementias (GARD) Cohort center. I am also greatly thankful to Prof. Kyu Yeong Choi, Prof. Jang Jae Lee and Prof. Jungsoo Gim for all the support, trainings, knowledge, time and the opportunity to work alongside.

I am also greatly thankful for the support of my colleagues Dr. Balaji Kannappan, Dr. Tamil Iniyan Gunasekaran, and Mr. Hojae Lim. Thanks to my other lab mates and administrative staffs for the much needed support throughout my thesis.

I would like to thank my family for always believing in me and supporting me silently. I would also like to thank my dearest Sehong Jeong. He has always stayed by my side, allowing me to do my research with ease.

I am also grateful to my classmates at Chosun University who have always supported me by my side. Because of them, I was able to overcome difficult situations without giving in.

I would like to express my sincere gratitude to all of you who have encouraged me. I will do my best to become a more essential person in the world by growing up in a more upright manner.