보건학 석사 학위논문

Implementing Alzheimer’s Disease Risk Prediction Model with Polygenic Risk Scores.

다유전자 위험 점수를 활용한

알츠하이머 질환 위험 예측 모델

2023년 02월

서울대학교 보건대학원

보건학과 보건학전공

임수진

Implementing Alzheimer’s Disease Risk Prediction Model with Polygenic Risk Scores.

지도 교수 원성호

이 논문을 보건학석사 학위논문으로 제출함
2022년 11월

서울대학교 보건대학원
보건학과 보건학전공
임수진

임수진의 보건학석사 학위논문을 인준함
2022년 12월

위 원 장 성주헌 (인)

부위원장 이우주 (인)

위 원 원성호 (인)

Abstract

Background

Alzheimer's Disease (AD), which accounts as the major cause of dementia, is highly influenced by genetic factors. Except for APOE, other common risk variants show moderate effect sizes, making it necessary to use polygenic genetic risk scores (PRSs) for predicting the genetic risk of AD. Because of genetic differences between populations, PRS need to be developed for each population. However, most of the AD PRS studies reported so far have been based on Europeans, and studies for East Asians have been limited to date, due to insufficient study samples which is critical for accurate genetic risk prediction. Therefore, research design to overcome the limitations of insufficient samples, such as the meta PRS method, which integrates multiple PRSs into one genetic risk score, is needed in predicting the genetic risk of AD.

Objective

This study aims to predict the genetic risk of AD in East Asians using the meta PRS method and compare the results with the PRS calculated through existing traditional methods.

Methods

Seven individual PRSs were derived by LDpred and 10-fold cross validation in the Korean Dementia Cohort(N=4,525). 7 PRSs are as follows: PRS using GWAS results of AD diagnosis status in Koreans (**KOR.AD**); PRS using GWAS summary statistics of AD diagnosis status in Europeans (**EUR.AD**); PRS using GWAS results of hippocampus (**Hippocampus**), amygdala volume (**Amygdala**) and entorhinal thickness (**Entorhinal**) in Koreans; and PRS using GWAS summary statistics of systolic blood pressure (**SBP**) and body mass index (**BMI**) in Japanese. 21 meta-PRS equations were derived and tested in Japanese dementia cohort(N=1,899) divided into validation set and test set. 21 meta-PRSs were derived by integrating individual PRSs using logistic regression on AD diagnosis status with ridge parameter and 10-fold cross-validation in the Japanese validation set. Test of meta-PRSs and original AD PRS (**KOR.AD**, and **EUR.AD**) was performed by measuring AD diagnosis status classification performances in the

Japanese test set.

Results

The best meta PRS, combination of **EUR.AD**, negative value of **Amygdala**, and **Entorhinal**, was validated to be significantly and positively associated with AD diagnosis status and the classification performance of the model including it was better (OR=1.277, 95% CI 1.15-1.41, AUC=0.698) than **KOR.AD** (OR=1.008, 95% CI 0.91-1.12, AUC=0.689) and **EUR.AD** (OR=1.257, 95% CI 1.14-1.39, AUC=0.695) in Japanese test set. Among 21 meta PRSs, meta PRSs including **EUR.AD** always showed bigger odds ratios with AD diagnosis status and the models with it always showed better classification performances than the others. If the PRSs of Alzheimer's Disease diagnosis status were conditioned negative value of **Amygdala** always showed bigger odds ratios with AD diagnosis status and improved classification performances.

Conclusion

This is the second study, which developed and tested AD PRS in East Asian sample, and in spite of the limited study samples, association of developed PRS with AD was significantly validated in the test set by using meta PRS method.

Keyword : Alzheimer's Disease, GWAS, PRS, Prediction model
Student Number : 2021-22216

Table of Contents

Chapter 1. Introduction **1**

Chapter 2. Methods **2**
2.1. Data Desciption 2
2.2. Statistical Analysis 5

Chapter 3. Results **8**
3.1. Genome Wide Association Study 8
3.2. Individual Polygenic Risk Scores 10
3.3. Meta Polygenic Risk Scores 10

Chapter 4. Discussion **12**

Reference **16**

Abstract **19**

Chapter 1. Introduction

As Alzheimer's Disease (AD) is highly influenced by genetic factors, with twin heritability estimated to be 58~79%, genome wide association studies (GWAS) have been actively conducted to identify related genetic factors[1]. Recent GWAS study of 790,000 Europeans identified 75 related locations[2]. Except for the ε4 allele of APOE, which shows large effect size on Alzheimer's Disease (odds ratio 3.7 in Europeans, and 5.7 in East Asians), other common risk variants show moderate effect sizes (odds ratio 0.86-1.22)[1, 2], which makes it difficult to find GWAS significant variants even with large sample sizes. This means that the Polygenic Risk Score(PRS), which adds together the influences of not only GWAS significant SNPs but also larger number of related SNPs, is suitable for genetic risk prediction of Alzheimer's Disease[3]. Predictive power of PRS, and its significant correlations with related biomarkers that previous studies have shown[4, 5] also support its usefulness.

Because of genetic differences between populations, PRS need to be developed for each population. However, most previous PRS studies of Alzheimer's Disease have been based on Europeans. In the case of East Asians, only one studies have been conducted on Chinese and case-control classification performance of models with PRS, even including APOE relevant SNPs, was relatively poor (AUC 0.61)[6]. This is because GWAS summary statistics derived from at least a hundred-thousand samples must be preceded for PRS development, and seven East Asian-based GWAS studies used a relatively small number of samples (1937-17031 East Asians), and their summary statistics are not disclosed [6-12]. Therefore, Alzheimer's Disease PRS studies for East Asians using larger samples are needed.

However, it is not easy to secure a sufficiently large number of East Asian samples. Recently, to overcome this limitation, meta PRS method, which

integrates multiple PRSs into one genetic risk score, have been proposed[13]. Meta PRS is different from conventional PRS methods, in the sense that it considers genetic variations that are not only directly related to the disease, but also those that are indirectly related, such as disease risk factors or various phenotypes in the course of the disease. In other words, meta PRS allows the complex consideration of heterogeneous genetic factors for each disease pathway, thus enhancing the statistical power of PRS.

Therefore, in this research, the meta PRS method was applied to predict the Alzheimer's Disease in East Asians. Prior to the development of meta PRS, GWAS were anlyzed for Alzheimer's Disease diagnosis status and MRI measurements for hippocampus, amygdala and entorhinal region in Koreans, because there is no released GWAS summary statistics. Then, using GWAS results and GWAS summary statistics released by previous studies, which were related to Alzheimer's Disease, seven individual PRSs were validated. By using various combinations of seven individual PRSs, meta PRSs were validated. Finally, best meta PRS was compared with original PRSs of Alzheimer's Disease. Flowchart of this study were shown in Figure 1.

Chapter 2. Methods

2.1. Data description

Study Samples

GWAS discovery and individual PRS validation dataset were based on Koreans over the age of 60, who have enrolled in the Gwangju Alzheimer's & Related Dementia (GARD) cohort registry at Chosun University in Gwangju, Korea. All subjects were evaluated by the dementia specialists in neurology and psychiatry at Chosun University Hospital and Chonnam National University Hospital, Gwangju, Republic of Korea. Subjects with no signs of neurological disorder and no problems with cognitive function or daily life performance were

classified as Cognitive Normal (CN), and subjects who met the NINCDS-ADRDA diagnostic criteria were classified as Alzheimer's Disease (AD) [14]. In addition to CN and AD group subjects, among the subjects who were classified as mild cognitive impairment (MCI) by NINCDS-ADRDA, only amnestic mild cognitive impairment (aMCI) patients were included as the subject for this study. Among them, 4525 Koreans (CN 2173, AD 1206 aMCI 1146) who passed the genetic data quality control criteria were included. After excluding aMCI patients, 3371 subjects (AD 1202, CN 2169) were analyzed for GWAS discovery and PRS validation of Alzheimer's Disease diagnosis status. This dataset was also used for PRS validation, using the European Alzheimer's Disease diagnosis status GWAS summary statistics derived from 111,326 clinically diagnosed AD cases and 677,663 controls[15]. A total of 5225 MRI data, including repeated measurements of 3291 subjects (AD 335, CN 1810, aMCI 1146), whose MRI test results are available, were analyzed for Alzheimer's Disease-related MRI phenotype GWAS discovery and PRS validation. 2,261 subjects who were measured for systolic blood pressure (SBP) and body mass index (BMI) were analyzed for SBP and BMI PRS validation using the GWAS summary statistics derived from 145,505 Japanese[16]. The meta PRS validation and test dataset was based on 1899 Japanese (CN 953, AD 946) with age over 60 for AD patients and over 70 for CN, who passed the quality control criteria of genetic data. Additional information about the test dataset can be found from previous study[8]. Summary statistics of study samples were shown in Table 1. Demographic information with the statistical significance was determined by t-test or Chi-square test, performed by Rex (Version 3.6.0, RexSoft Inc., Seoul, Korea) [17].

Ethincal considerations

The Institutional Review Boards of the participating hospitals (Chosun University Hospital:2-1041055-AB-N-01-2020-37/2021-58/ 2022-29/ 2022-53/ 2023-04) approved the study protocol and waived additional informed consent of patient

participation. The participants provided written informed consent and participation was voluntary.

Genotyping, quality control, imputation, and principal component analysis procedures

4,525 Korean cohort subjects were genotyped with Affymetrix customized Korean Chip version1.0, and version1.1[18, 19]. 1899 Japanese cohort subjects were genotyped with Affymetrix GeneChip 6.0[8]. Genotype datasets were preprocessed, using PLINK[20], and standard downstream quality control and imputation were performed. For quality control, SNPs, with low genotype call rate (< 95%), and those which significantly deviated from Hardy-Weinberg equilibrium($p < 1\times10^{-5}$), were excluded[21]. In the case of Korean cohort, in order to remove the batch effect of different chip versions, the association of each SNP with the chip version (1.0 or 1.1) was estimated using logistic regression analysis. SNPs which were significantly different ($p < 1\times10^{-5}$) between chip versions were also excluded. And individuals with low genotype call rate (< 95%), sex inconsistency, heterozygosity rate greater than three standard deviations from the average heterozygosity rate, and those that presented cryptic relatedness, or considered outlier (out of 5*IQR range) by analysis of principal components (PC) were excluded. Quality controlled SNPs were imputed to improve genotyping coverage. In the case of Korean cohort, pre-phased reference haplotypes from the Haplotype Reference Consortium (HRC) panel version 1.1 of the Michigan Imputation server was used[22], and the Japanese cohort used the Northeast Asian Reference Database (NARD) panel of the NARD Imputation Server[23]. Imputed SNPs were excluded additionally by genotype call rate (< 95%), Hardy-Weinberg equilibrium ($p < 1\times10^{-6}$), Minor Allele Frequency (< 0.01), and imputation quality info score (<0.5). Finally, 6625043 SNPs were left in the Korean AD diagnosis data, 6629747 SNPs in the Korean MRI phenotype data, and 7081334 SNPs in the Japanese data.

Brain MRI acquisition and processing

Of the total 5225 Brain MRI measurement data, 5098 were measured using 3.0 T MRI (Skyra, Siemens, TR = 2300 ms; TE = 2.143 ms; TI = 900 ms; 9 flip angle; FoV = 256x256; matrix = 320x320; number of slices = 178), and 127 were measured using 1.5 T MRI (n=125) scanners (Magnetom Avanto, Siemens, TR = 1800 ms; TE = 3.43 ms; TI = 1100 ms; 15 flip angle; FoV = 224x224; matrix = 256x256; number of slices = 176). In order to remove the batch effect due to the different measurement methods, MRI phenotypes were standardized for each subgroup (3.0T MRI measured group&1.5T MRI measured group), and MRI Tera information was additionally used as a covariate in GWAS analysis and PRS equation development. The Alzheimer's Disease-related phenotypes used in this study were Hippocampus and amygdala volume and entorhinal thickness, which were reported to be highly related to AD [4, 5, 24].

2.2. Statistical Analysis

Genome Wide Association Study

GWAS for Alzheimer's Disease diagnosis status and hippocampus, amygdala volume, and entorhinal thickness, which do not have a released GWAS summary statistics, were analyzed. The association of genotype dosages of each SNP with the AD diagnosis status was estimated by logistic regression analysis adjusting for age, gender, first 10 PC components, and with or without APOE ε4 allele carrier status. PLINK software was used to analyze the association and calculate the principal components[20]. The associations of genotype dosages of each SNP with the hippocampus, amygdala volume, and entorhinal thickness were estimated by the linear mixed regression to account for the random effects on individuals in repeated measurements. All MRI measurements used in the analysis were standardized as mentioned above. Linear mixed regressions were adjusted for age, gender, first 10 PC components, MRI Tera as a two-level

categorical value (1.5T, and 3.0 T), scaled ICV and with or without APOE ε4 allele carrier status. In this analysis, R package LmerTest was used [25]. Visualization of the GWAS result and functional annotation, including that of genes mapping to the identified risk loci, were conducted using the FUMA web application (https://fuma.ctglab.nl/)[26].

Individual PRS derivation

Seven individual PRSs were developed with subjects of the Korean cohort using four GWAS results conducted in this study, and three previously released GWAS summary statistics. The seven PRSs were divided into three types, which are PRSs of Alzheimer's Disease diagnosis status, related MRI measurements, and related risk factors. **KOR.AD** and **EUR.AD** are PRSs of Alzheimer's Disease diagnosis status which were derived from Korean based GWAS and European based GWAS each. They were original PRSs for Alzheimer's Disease. **Hippocampus, Amygdala,** and **Entorhinal** are PRSs of related MRI measurement developed with Korean Hippocampus volume GWAS, Korean Amygdala volume GWAS, and Korean Entorhinal thickness GWAS each. **SBP** and **BMI** are two PRSs of related risk factors, developed with Japanese systolic blood pressure GWAS, and Japanese body mass index GWAS each. In order to estimate the additional predictive power of SNPs excluding the APOE-related SNPs, SNPs in the same LD with APOE region (SNPs r2 >0.5 with rs7412 and rs429358) were excluded for the derivation of **KOR.AD, EUR.AD, Hippocampus, Amygdala, and Entorhinal**. In the case of related MRI measurement PRSs **(Hippocampus, Amygdala, Entorhinal),** the smaller the PRS value, the higher the correlation with Alzheimer's Disease due to the atrophy of the corresponding region. Thus, they were used as negative values for meta PRS analysis later on.

Each individual PRS was calculated by LDpred2, which re-estimates the distribution of effect size accounting for the LD structure using Bayesian

approaches. R package *bigsnpr* was used for analysis [27]. In-sample LD was calculated using the function provided in the package, and heritability was also estimated through the built-in ldsc regression function. The grid method was used to estimate the parameters. Multiple values for each parameter were tested (Proportion of causal SNPs (p): 1, 0.03, 0.01, 0.003, 0.001, SNP heritability (h^2): h^2 estimate $\times$ (0.7, 1, 1.4)). The performances of 15 parameter combinations were evaluated. Among the 15 parameter combinations, the PRS equation showing either the lowest mean MAE or highest mean AUC was selected as the best model through 10-fold cross validation. Prediction model with **KOR.AD** or **EUR.AD**, adjusted for age, gender, 10 PC components, and APOE ε4 allele carrier status, predicted Alzheimer's Disease diagnosis status by logistic regression. Prediction model with **Hippocampus, Amygdala,** or **Entorhinal**, adjusted for age, gender, 10 PC components, scaled ICV, MRI Tera information as a two-level categorical variable (1.5T, and 3.0T) and APOE ε4 allele carrier status, predicted each MRI measurements (scaled hippocampus, amygdala, and entorhinal measurements) by linear regression. Prediction model with **SBP** or **BMI**, adjusted for age, gender, and 10 PC components, predicted each target phenotype (SBP, and BMI) by linear regression.

Prior to the calculation of meta PRS, Pearson correlation between each PRS and Alzheimer's Disease diagnosis status was analyzed. As a result, although higher systolic blood pressure and higher body mass index are risk factors, PRSs of them (**SBP** and **BMI**) showed negative correlation coefficients with Alzheimer's Disease diagnosis status, and were consequently excluded from the derivation of meta PRS. Therefore, five PRSs (**KOR.AD, EUR.AD, and negative value of Hippocampus, Amygdala, and Entorhinal**) were used to derive meta PRS.

meta PRS derivation and test

Japanese cohort data was divided into two for development and testing of

meta PRS. The meta PRS was derived by revising the method introduced by Grad Abraham, et.al[13]. In short, each PRS was standardized (unit standard deviation) and association of each PRS with AD status was estimated by logistic regression with ridge penalty parameter adjusted for gender, 10 PC components, and APOE ε4 allele carrier status, using the R package 'glmnet'[28]. The ridge penalty parameter was applied because of the significant correlations between individual PRSs. Age was not adjusted because age limit was placed on the recruitment of case and control groups in Japanese cohort. To compare the performance of meta PRSs with original PRSs (**KOR.AD and EUR.AD**), total 21 meta PRSs, combinations of at least one original AD PRS and at least one related MRI measurement PRS, were developed. Ridge penalty parameter was optimized by highest 10-fold cross validated AUC.

The final estimated coefficients for each PRS (log odds ratios) were fixed for test. The final per-PRS log odds were $\gamma_1 \cdots \gamma_k$ converted to an equivalent per SNP score via a weighted sum.

$$\text{Meta PRS} \propto \sum_{j=1}^{m} x_{ij} \left(\frac{\gamma_1}{\sigma_1}\alpha_{j1} + \cdots + \frac{\gamma_k}{\sigma_k}\alpha_{jk}\right)$$

where m is the total number of SNPs, k is the index of individual PRSs (k=1, … ,5), $\sigma_1 \cdots \sigma_k$ are the empirical standard deviations of each PRS in the derivation data, $\alpha_{j1} \cdots \alpha_{jk}$ are the estimated SNP effect sizes from the LDpred for the jth SNP in each PRS, and x_{ij} is genotype (0,1,2) for the ith individual's jth SNP. By using fixed coefficients, 21 meta PRSs were calculated in an independent test set to compare the odds ratio of each meta PRS and confirm the predictive performance.

Chapter 3. Results

3.1. Genome Wide Association Study

Alzheimer's Disease Status

As a the result of the Alzheimer's Disease diagnosis status GWAS, which was not adjusted for APOE ε4 carrier status, there were 53 genome-wide significant ($p<5\times10^{-8}$) SNPs across four genes, namely APOE($p= 1.76\times e^{-37}$), PVRL2($p= 7.98\times e^{-28}$), TOMM40($p= 1.42\times e^{-37}$), and APOC($p= 4.28\times e^{-33}$) (genomic inflation=1.03), which were found to be associated with Alzheimer's Disease in the previous Korean sample GWAS studies[29]. As a result of the adjusted APOE ε4 carrier status GWAS, there were only two genome-wide significant SNPs across the APOE ($p= 3.57\times e^{-11}$) and TOMM40($p= 5.16\times e^{-11}$) gene (genomic inflation=1.033) (Table2). The QQ plot and Manhattan plot for the GWAS analysis can be found in Figure 2.

Brain MRI measurements: hippocampus, amygdala volume and entorhinal thickness

As a result of the hippocampus, amygdala volume and entorhinal thickness GWAS, which was not adjusted for APOE ε4 carrier status, there was only one genome-wide significant ($p<5\times10^{-8}$) SNP across the SHARPIN gene (rs77359862; 8:145154282:A:G). The p-values and genomic inflation factors were $p= 2.37\times e^{-8}$, genomic inflation= 1.014 for hippocampus, $p= 2.21\times e^{-10}$, genomic inflation= 1.006 for amygdala, and $p= 1.68\times e^{-13}$, genomic inflation= 1.008 for entorhinal. Among the four genes, which were associated with AD status, only the APOE and TOMM40 genes showed association ($p<1.0\times e^{-6}$) with hippocampus volume and entorhinal thickness. While a previous study also reported a significant APOE gene association ($p<5.0\times e^{-8}$), results of this study showed a stronger association[24]. Even after adjusting for APOE ε4 carrier status, SNP across SHARPIN gene (rs77359862; 8:145154282:A:G) was GWAS significant with all three regions (hippocampus $p= 2.2\times e^{-8}$, genomic inflation= 1.014, amygdala $p= 2.15\times e^{-10}$, genomic inflation= 1.003, entorhinal $p= 1.60\times e^{-13}$, genomic inflation= 1.007). GWAS significant association of SHARPIN gene with hippocampus volume and entorhinal thickness even after adjusting for

APOE ε4 carrier status is consistent with the results of previous studies. Moreover, significant association was also shown for amygdala region, which had not been identified previous studies (Table3). The QQ plot and Manhattan plot for the GWAS analysis can be found in Figure 2.

3.2. Individual Polygenic Risk Scores

After derive seven individual PRSs, the correlations between seven individual PRSs and Alzheimer's Disease diagnosis status in the Korean and Japanese set were shown in Figure 3. Significant high correlations between 3 related MRI measurement PRSs (**hippocampus**, **amygdala** and **entorhinal**) were confirmed in independent Japanese dataset. Pearson correlation of **Hippocampus** with **Amygdala** was 0.71, **Hippocampus** with **Entorhinal** was 0.55, and **Amygdala** with **Entorhinal** was 0.53. The significant correlations of 3 related MRI measurements PRSs with **KOR.AD** was also found in independent Japanese dataset. Negative value of **Entorhinal** showed highest correlation with **KOR.AD** among them (Pearson correlation 0.23). For the correlation between 7 individual PRSs and AD diagnosis status, in the case of **SBP** and **BMI**, even in the Korean validation set, they showed negative correlations with Alzheimer's disease diagnosis status. For that reason, they were excluded from the meta-PRS derivation. The correlations between Alzheimer's Disease diagnosis status and the remaining 5 PRSs were significant as positive correlation in validation set, but only **EUR.AD** was significant in independent Japanese set (Pearson correlation 0.09). That difference may be due to genetic differences between Koreans and Japanese, but overfitting cannot be ruled out. In the case of 4 PRSs using Korean GWAS results, all samples used for the PRS derivation were also included in the GWAS discovery set. Therefore, despite the application of the 10-fold cross validation, there still remains risk of overfitting. That's the reason why validation and testing of meta-PRS used split independent Japanese set.

3.3. Meta Polygenic Risk Scores

Associations of each meta PRS per standard deviation with Alzheimer's Disease diagnosis status were estimated by logistic regression, adjusted for gender, 10 PC components, and APOE ε4 carrier status in the Japanese test set (Table 4, and Figure 4). In the case of original AD PRSs, the association of **KOR.AD** was not significant (OR=1.008, 95% CI 0.91-1.12), whereas **EUR.AD** was significant (OR=1.257, 95% CI 1.14-1.39). The meta PRS of **EUR.AD,** negative value of **Hippocampus, Amygdala,** and **Entorhinal** and the meta PRS of **KOR.AD**, **EUR.AD,** negative value of **Hippocampus, Amygdala,** and **Entorhinal** showed the strongest association (OR=1.285, 95% CI 1.16-1.42). Odds ratio of meta PRS (**EUR.AD**, negative value of **Amygdala** and **Entorhinal**), which was included in best prediction model, was also significant (OR=1.277, 95% CI 1.15-1.41). The meta PRSs including **EUR.AD** always showed significant associations with Alzheimer's Disease diagnosis status, and their odds ratios and confidence intervals did not show statistically significant differences between each other. If the PRSs of Alzheimer's Disease diagnosis status were conditioned, in other words, meta PRSs including **KOR.AD**, meta PRSs including **EUR.AD**, or meta PRSs including **both** were compared within their condition, including negative value of **Amygdala** additionally always increased association with Alzheimer's Disease diagnosis status.

Alzheimer's Disease diagnosis status classification performances of the models including each meta PRS, gender, 10 PC components, and APOE ε4 carrier status were tested by logistic regression in the Japanese test set (Table 4, Figure 5). The area under the curve (AUC) of null model including only gender, 10 PC components was 0.518, and the AUC of APOE ε4 model including APOE ε4 carrier status in addition to the null model was 0.690. In the case of original AD PRSs, AUC of the model with **KOR.AD** was even smaller than the APOE ε4

model (AUC=0.689), whereas **EUR.AD** improved performance (AUC=0.695). The model with meta PRS of **EUR.AD,** negative value of **Amygdala,** and **Entorhinal** showed the best performance (AUC=0.698). The models with meta PRSs including **EUR.AD** always improved performances than the others. If the PRSs of Alzheimer's Disease diagnosis status were conditioned, likewise the association, the models including negative value of **Amygdala** additionally always showed higher classification performances. The comparison of AUC-ROC Curve between null model, APOE ε4 model, and best meta PRS model can be found in Figure 6.

Chapter 4. Discussion

In this paper, meta PRSs of Alzheimer's Disease were derived and compared with original Alzheimer's Diseases PRSs (**KOR.AD,** and **EUR.AD**). The best meta PRS, combination of **EUR.AD**, negative value of **Amygdala**, and **Entorhinal**, was validated to be significantly and positively associated with Alzheimer's Disease diagnosis status and the classification performance of the model including it was better (OR=1.277, 95% CI 1.15-1.41, AUC=0.698) than **KOR.AD (**OR=1.008, 95% CI 0.91-1.12, AUC=0.689) and **EUR.AD (**OR=1.257, 95% CI 1.14-1.39, AUC=0.695**)** in Japanese test set.

Among 21 meta PRSs, meta PRSs including **EUR.AD** always showed bigger odds ratios with Alzheimer's Disease diagnosis status and the models with it always showed better classification performances than the others. It is an interesting result that PRS using European GWAS summary statistics was more useful than PRS using Korean GWAS results in predicting genetic risks for Alzheimer's Disease in East Asian. The possible reason why **KOR.AD** did not perform well is that the sample size of the GWAS discovery set, which is critical for genetic risk prediction of polygenic disease was extremely small (N=3,371) compared to European GWAS summary statistics (N=788,989), and moreover

there was sample overlap between GWAS discovery set and PRS validation set. On the contrary, in the case of **EUR.AD**, the effect size of each SNP in the GWAS summary statistics was re-estimated reflecting the LD structure of Koreans using the LD-pred method for PRS calculation, which could improve the prediction performance in East Asian.

The other interesting result is that if the PRSs of Alzheimer's Disease diagnosis status were conditioned negative value of **Amygdala** always showed bigger odds ratios with Alzheimer's Disease diagnosis status and improved classification performances. This is interesting because hippocampus volume is known as one of the most relevant regions to Alzheimer's Disease[30]. In this study, hippocampus volume also showed larger Pearson correlation(r=-0.508 , 95% CI [-0.539, -0.476]) with Alzheimer's Disease diagnosis status than amygdala volume (r=-0.413, 95%CI [-0.448, -0.377]), and negative value of hippocampus PRS also showed larger odds ratio per standard deviation (OR=3.3439, 95%CI [3.101, 4.741]) than negative value of amygdala PRS (OR=3.146, 95%CI [2.572, 3.850]) in Korean validation set. However, in Japanese test set, among related MRI measurement PRSs, nothing showed significant associations with Alzheimer's Disease diagnosis (**Hippocampus** OR=1.003, 95%CI [0.907, 1.108], **Amygdala** OR=1.065, 95%CI [0.963, 1.177], **Entorhinal** OR=0.996, 95%CI [0.901, 1.101]). This result may be due to genetic differences between Koreans and Japanese[31]. According to Miyashita, A., et al.(2022), there was no common genetic variant, except APOE region among 7 East Asian Alzheimer's Disease GWAS researches[32]. Therefore, the larger and various East Asian samples are required for generalization of PRS in East Asian and it is necessary to validate this study in an independent Korean sample in the future.

In addition to the key discovered facts above, this study has several importance. The first is that, this is the second study, which developed and tested Alzheimer's Disease PRS in East Asian sample. Zhou, X., et al.(2020) developed

PRS equation using 41 SNPs with Chinese population(n=729) and tested PRS with Chinese population(n=402)[6]. This study, however, have a more generalizability of the result because of using larger sample size, more various genetic variants, and being tested with samples of different nationalities.

The second importance of this research is that in spite of the small GWAS sample size, association of developed PRS with Alzheimer's Disease diagnosis status was significantly validated in the test set by using meta PRS method. It means that the problem of lack of sample size in **KOR.AD** and the problem of validation in **EUR.AD** due to genetic difference between each ethnicity are both overcame in a sense, by using meta PRS method. Although the minimum number of GWAS samples required for validation of PRS is known as 100,000 or more, significant results were reported using a much smaller sample size, suggesting that more meaningful research results can be obtained by this research approach after securing a sufficient number of samples.

Lastly, this research complemented the limitation of binary classification of AD diagnosis by integrating not only directly Alzheimer's Disease relevant genetic variants but also indirectly relevant genetic variants. Case/Control binary classification includes the possibility of incorrect diagnosis, and errors that were classified as control group at the time of sample collection, but will become a case group in the future. Alzheimer's Disease is more fatal to these errors because its incidence rate is critically increased due to aging, and it is diagnosed based on cognitive decline, which appears more later than atrophy of the brain structure[33].

In addition, to alleviate misclassification errors, most studies exclude mild cognitive impairment patients from the study, resulting in loss of samples. However, by including PRSs for related MRI phenotypes in this study, the statistical power is greater, and the possibility of future outbreaks is relatively reflected[34].

This study has several limitations. First of all, although meta PRS showed significant association, the degree of improvement in predictive power was small, so it did not have the meaning applicable in clinical setting. A likely reason is that, firstly east Asian GWAS sample sizes are still limited. Secondly due to sample overlap between GWAS discovery set and PRS derivation set, overfitting could make true best model be selected. Thirdly, we consider only seven PRSs for meta PRS derivation, because of data restriction.

Therefore, I expect East Asian Alzheimer's Disease GWAS progress and independent PRS derivation set can be used, then this meta PRS can be more powerful. And PRSs of Alzheimer's Disease related risk factors like smoking and coronary artery disease and related phenotype like cognitive function score, PET measurements, and CSF measurements can be used for more predictive meta PRS development.

Table 1. Summary statistics

		Korean populations (Number of Samples=4525)				Japanese populations (Number of Samples =1899)		
variables	Statistics	CN N=2173	AD N=1206	MCI N=1146	P-value	CN N=953	AD N=946	P-value
All samples								
samples	N	2173	1206	1146		953	946	
Age at exam	mean ± Std	73.48 ± 5.73	74.83 ± 6.83	73.99 ± 5.99	<0.001	77.00 ± 5.90	72.99 ± 4.26	<0.001
Female	N (%)	1339 (61.62%)	749 (62.11%)	636 (55.50%)	<0.001	550 (57.71%)	681 (71.99%)	<0.001
APOE ε4 allele non-carrier								
samples	N	1608	679	839		794	418	
Age at exam	mean ± Std	73.92 ± 5.70	75.53 ± 6.93	74.13 ± 6.00	<0.001	77.18 ± 5.98	73.36 ± 4.30	<0.001
Female	N (%)	991 (61.63%)	413 (60.82%)	452 (53.87%)	<0.001	459 (57.81%)	300 (71.77%)	<0.001
APOE ε4 allele carrier								
samples	N	564	526	307		159	528	
Age at exam	mean ± Std	72.25 ± 5.59	73.92 ± 6.61	73.63 ± 5.96	<0.001	76.08 ± 5.38	72.70 ± 4.22	<0.001
Female	N (%)	347 (61.52%)	335 (63.69%)	184 (59.93%)	0.5364	91 (57.23%)	381 (72.16%)	<0.001

(Abbreviation)CN, cognitive normal; AD, Alzheimer's disease; MCI, Mild Cognitive Impairment; Std, standard deviation

Table 2. Significant GWAS results ($p < 1.0 \times 10^{-7}$) for Alzheimer's Disease diagnosis status

chr	pos	non effect allele	effect allele	MAF	P-value	Beta	Std.Error	Nearest gene
not adjusted for APOE ε4 carrier								
19	45406673	G	A	0.09623	1.42E-37	2.406	0.06853	TOMM40
19	45411941	T	C	0.08631	1.76E-37	2.443	0.06983	APOE
19	45421254	G	A	0.09722	4.28E-33	2.312	0.06993	APOC1
19	45387459	C	G	0.09524	7.98E-28	2.201	0.07216	PVRL2
adjusted for APOE ε4 carrier								
19	45411941	T	C	0.08631	3.57E-11	4.435	0.225	APOE
19	45406673	G	A	0.09623	5.16E-11	2.852	0.1596	TOMM40

(Abbreviation) chr, chromosome; MAF, minor allele frequency; Std.Error, standard error

Table 3. Significant GWAS results ($p < 1.0 \times 10^{-7}$) for MRI measurements related to Alzheimer's Disease.

chr	pos	non effect allele	effect allele	MAF	P-value	Beta	Std.Error	Nearest gene
Hippocampus volume; not adjusted for APOE e4 carrier								
8	145154282	G	A	0.03968	2.20E-08	-0.428	0.07622	SHARPIN
Hippocampus volume; adjusted for APOE e4 carrier								
8	145154282	G	A	0.03968	2.37E-08	-0.428	0.07640	SHARPIN
Amygdala volume; not adjusted for APOE e4 carrier								
8	145154282	G	A	0.03968	2.15E-10	-0.506	0.07946	SHARPIN
Amygdala volume; adjusted for APOE e4 carrier								
8	145154282	G	A	0.03968	2.21E-10	-0.507	0.07950	SHARPIN
Entorhinal thickness; not adjusted for APOE e4 carrier								
8	145154282	G	A	0.03968	1.60E-13	-0.638	0.08603	SHARPIN
Entorhinal thickness; adjusted for APOE e4 carrier								
8	145154282	G	A	0.03968	1.68E-13	-0.638	0.08622	SHARPIN

(Abbreviation) chr, chromosome; MAF, minor allele frequency; SE, standard error

Table 4. Associations of each meta PRS with Alzheimer's Disease diagnosis status and classification performances of models including each meta PRS in Japanese test set

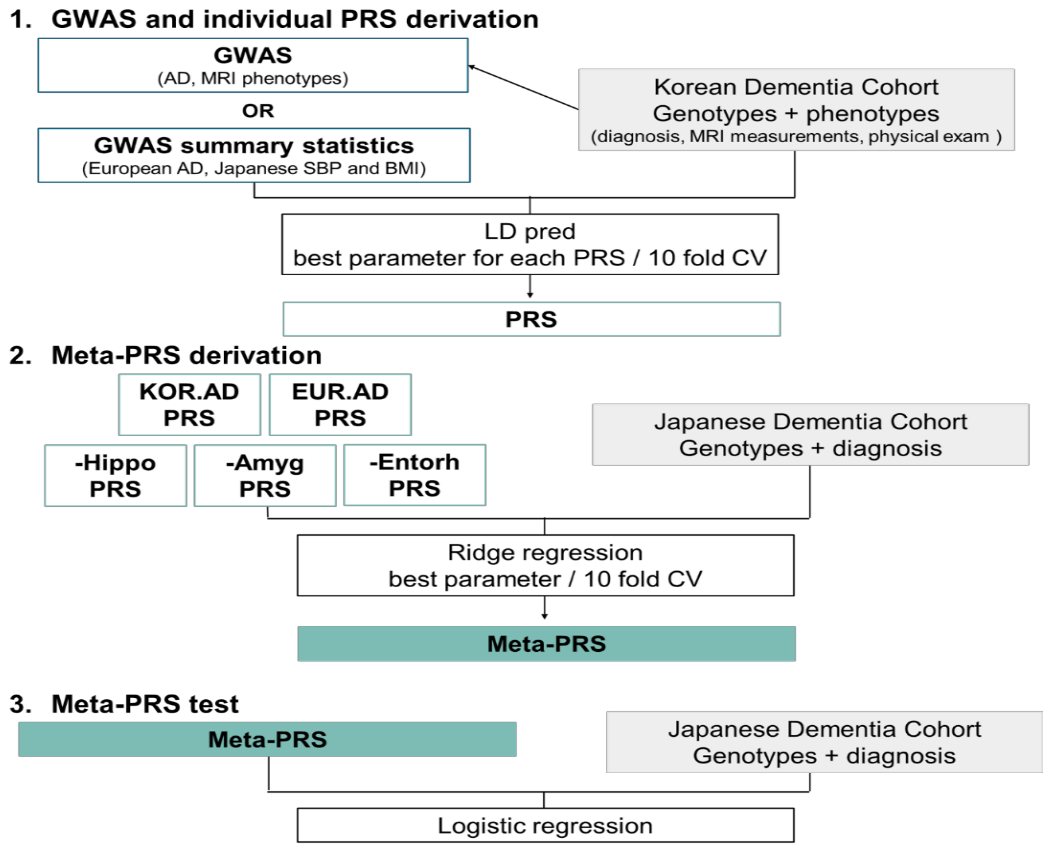
meta PRSs	ln OR	SE	z value	P-value	AUC
Meta PRSs with KOR.AD					
KOR.AD	0.008	0.052	0.146	0.884	0.689
KOR.AD+(-Hippocampus)	0.007	0.052	0.134	0.894	0.688
KOR.AD+(-Amygdala)	0.061	0.051	1.187	0.235	0.690
KOR.AD+(-Entorhinal)	0.005	0.052	0.096	0.924	0.689
KOR.AD+(-Hippocampus)+(-Amygdala)	0.087	0.051	1.692	0.091	0.693
KOR.AD+(-Hippocampus)+(-Entorhinal)	0.003	0.052	0.067	0.947	0.688
KOR.AD+(-Amygdala)+(-Entorhinal)	0.076	0.051	1.477	0.140	0.693
KOR.AD	0.090	0.051	1.759	0.079	0.693

+(-Hippocampus)+(-Amygdala)+(-Entorhinal)					
Meta PRSs with EUR.AD					
EUR.AD	0.229	0.052	4.408	1.04E-05	0.695
EUR.AD+(-Hippocampus)	0.229	0.052	4.406	1.05E-05	0.695
EUR.AD+(-Amygdala)	0.240	0.052	4.612	3.99E-06	0.696
EUR.AD+(-Entorhinal)	0.228	0.052	4.401	1.08E-05	0.694
EUR.AD+(-Hippocampus)+(-Amygdala)	0.250	0.052	4.788	1.68E-06	0.697
EUR.AD+(-Hippocampus)+(-Entorhinal)	0.228	0.052	4.398	1.09E-05	0.694
EUR.AD+(-Amygdala)+(-Entorhinal)	0.244	0.052	4.693	2.70E-06	0.698
EUR.AD +(-Hippocampus)+(-Amygdala)+(-Entorhinal)	0.251	0.052	4.814	1.48E-06	0.697
Meta PRSs with KOR.AD and EUR.AD					
KOR.AD+EUR.AD	0.228	0.052	4.406	1.05E-05	0.693
KOR.AD+EUR.AD+(-Hippocampus)	0.228	0.052	4.407	1.05E-05	0.693
KOR.AD+EUR.AD+(-Amygdala)	0.239	0.052	4.604	4.15E-06	0.697
KOR.AD+EUR.AD+(-Entorhinal)	0.228	0.052	4.404	1.07E-05	0.693
KOR.AD+EUR.AD +(-Hippocampus)+(-Amygdala)	0.249	0.052	4.783	1.73E-06	0.697
KOR.AD+EUR.AD +(-Hippocampus)+(-Entorhinal)	0.228	0.052	4.400	1.08E-05	0.693
KOR.AD+EUR.AD +(-Amygdala)+(-Entorhinal)	0.244	0.052	4.689	2.75E-06	0.698
KOR.AD+EUR.AD +(-Hippocampus)+(-Amygdala)+(-Entorhinal)	0.251	0.052	4.813	1.48E-06	0.698

The associations of each meta PRS per standard deviation with Alzheimer's Disease diagnosis status were estimated by logistic regression, adjusted for gender, 10 PC components, and APOE ε4 carrier status in the Japanese test set. The Alzheimer's Disease diagnosis classification performances of the models including each meta PRS

were measured by AUC. The AUC of APOE ε4 model included only APOE ε4 allele carrier status, gender, and 10 PCs, was 0.690. (Abbreviation) **KOR.AD**, AD PRS derived with Korean AD GWAS; **EUR.AD**, AD PRS derived with European AD GWAS; **(-Hippocampus)**, negative value of hippocampus volume PRS derived with Korean hippocampus GWAS; **(-Amygdala)**, negative value of amygdala volume PRS derived with Korean amygdala GWAS; **(-Entorhinal)**, negative value of entorhinal thickness PRS derived with Korean entorhinal GWAS; ln OR, log value of odds ratio; SE, standard error; AUC, area under the curve;

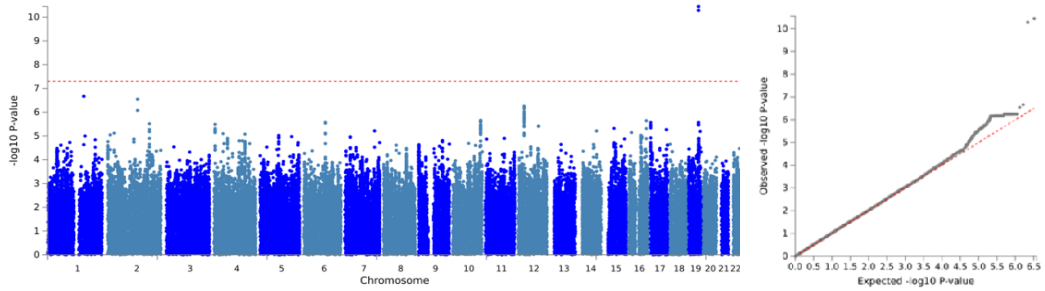
Figure 1. Flow chart of the study



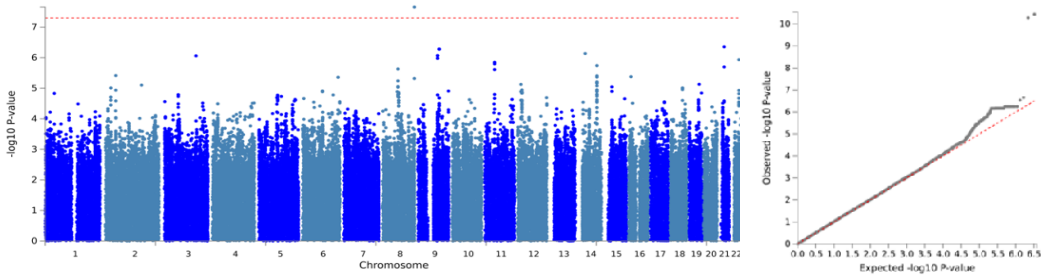
1. Individual PRSs were derived by LDpred and 10 fold cross validation in the Korean Dementia Cohort using the results of GWAS for four individual traits and GWAS summary statistics for three individual traits. **2.** The meta-PRS for Alzheimer's Disease was then derived by integrating individual PRSs using Ridge regression and 10 fold cross-validation in the Japanese validation set. **3.** Test of the meta-PRS for Alzheimer's Disease was performed in the Japanese test set. (Abbreviation) GWAS**, genome-wide association study;** PRS**, polygenic risk score;** AD**, Alzheimer's Disease;** MRI phenotypes**, hippocampus, amygdala volume and entorhinal thickness;** European AD, **Alzheimer's Disease GWAS summary statistics based on European population;** Japanese SBP and BMI**, systolic blood pressure and body mass index GWAS summary statistics based on Japanese population; KOR.AD PRS**, AD PRS derived with Korean AD GWAS; **EUR.AD PRS**, AD PRS derived with European AD GWAS; **-Hippo PRS,** negative value of hippocampus volume PRS derived with Korean hippocampus GWAS; **-Amyg PRS**, negative value of amygdala volume PRS derived with Korean amygdala GWAS; **-Entorh PRS**, negative value of entorhinal thickness PRS derived with Korean entorhinal GWAS; **CV,** cross validation.

Figure 2. Manhattan plot and QQ plot of APOE ε4 allele carrier status adjusted GWAS

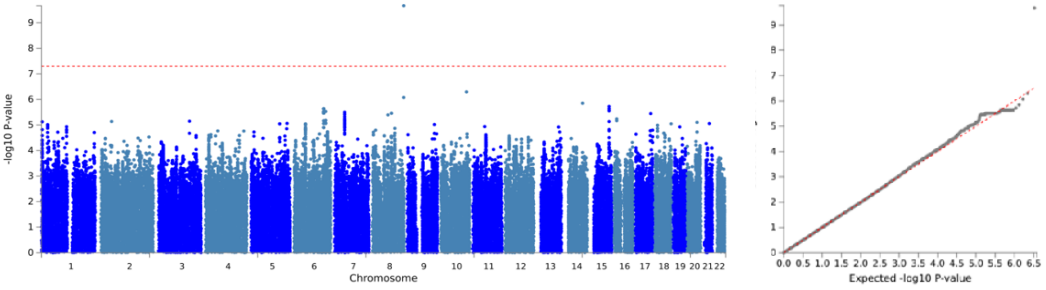
1. Alzheimer's Disease diagnosis status GWAS



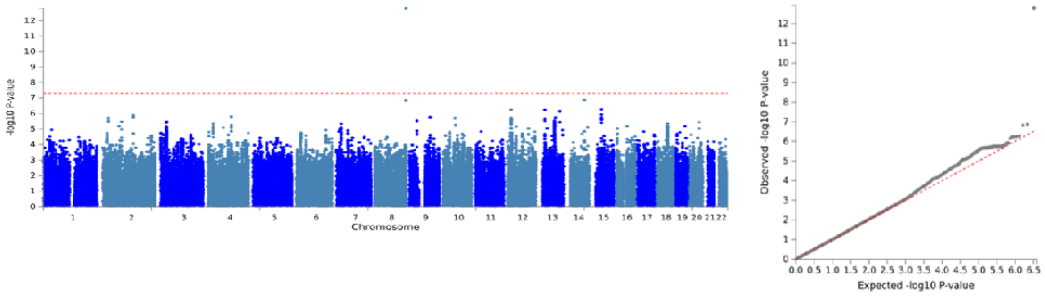
2. Hippocampus volume GWAS



3. Amygdala volume GWAS



4. Entorhinal thickness GWAS

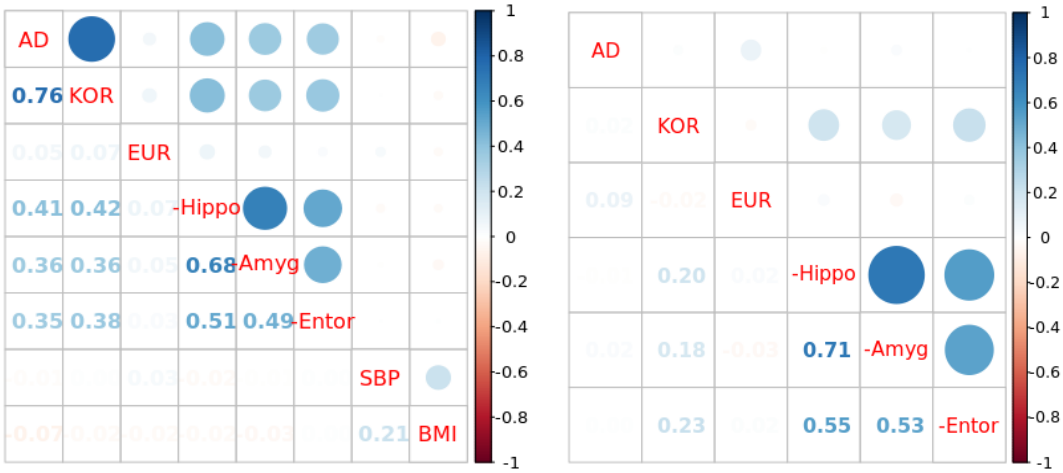


1. Alzheimer's Disease diagnosis status GWAS**,** adjusted for APOE ε4 allele carrier status, age, gender, and 10 PC components, was derived in the 3371 Korean Dementia

Cohort samples (Alzheimer's Disease 1202, and Cognitive Normal 2169) by logistic regression. Inflation factor lambda was 1.033. **2~4.** MRI measurements related to Alzheimer's Disease (Hippocampus, Amygdala volume and Entorhinal thickness) GWAS**,** adjusted for APOE ε4 allele carrier status, age, gender, 10 PC components, MRI Tera, Intracranial volume and random effect of individual were derived in the 5225 Korean Dementia Cohort MRI observations (Alzheimer's Disease 335, Cognitive Normal 1810, and amnestic mild cognitive impairment 1146) by linear mixed model. Inflation factors were 1.014, 1.003, and 1.007.

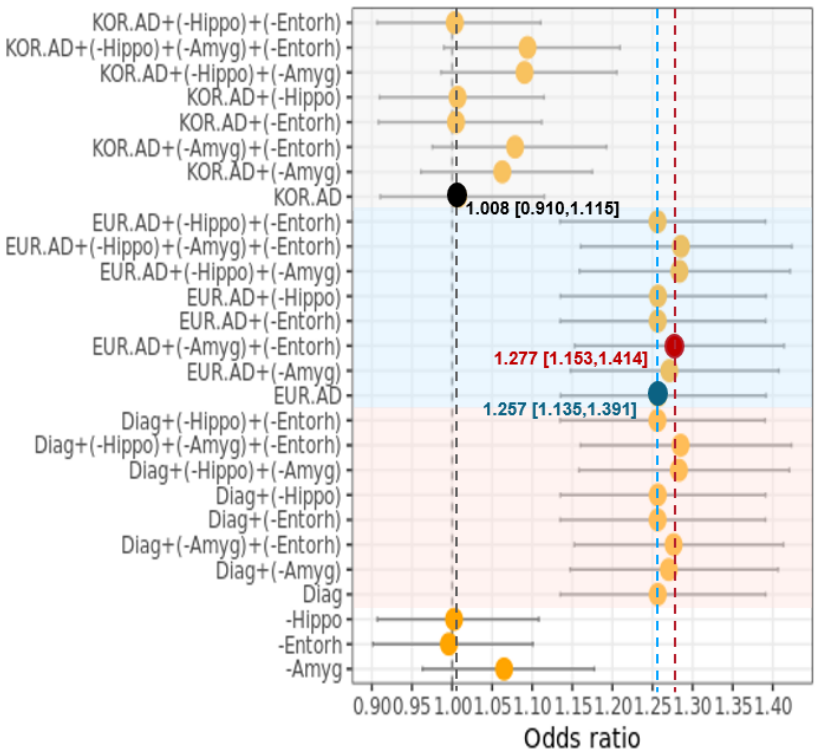
Figure 3. Correlation matrix of 7 individual PRSs and AD diagnosis status.

1. Individual PRSs Validation set(Korean) **2. Independent Japanese set**



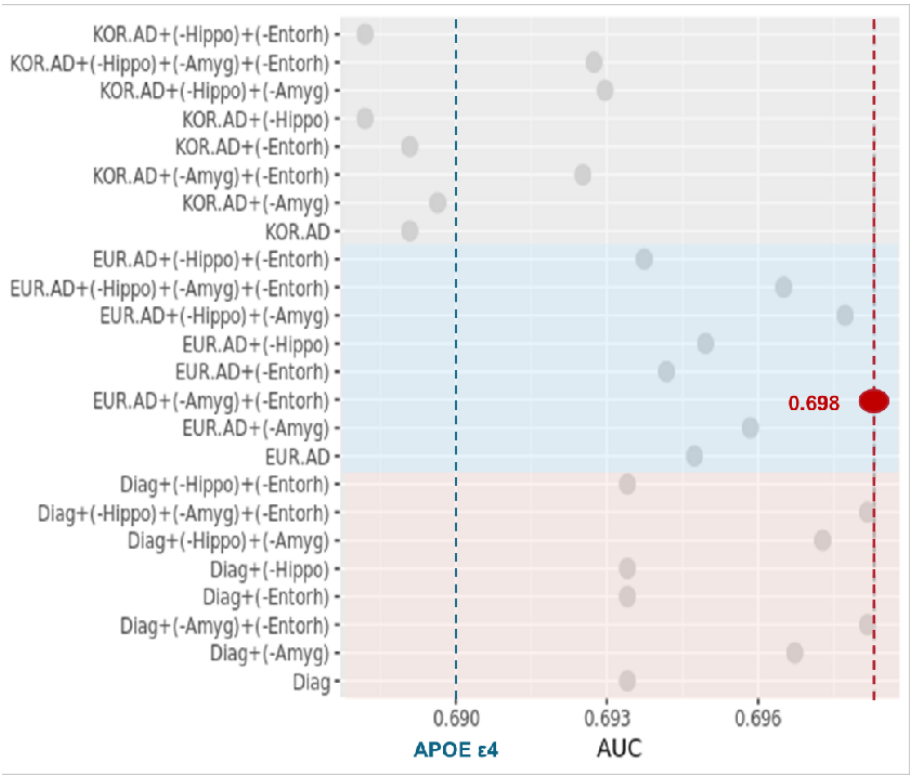
1. shown is the Pearson correlation of individual PRSs in a PRS validation sample of 1366 Koreans.
2. shown is the Pearson correlation of individual PRSs in an independent sample of 1899 Japanese.
(Abbreviation) **AD**, Alzheimer's disease diagnosis status; **KOR**, AD PRS derived with Korean AD GWAS; **EUR**, AD PRS derived with European AD GWAS; **-Hippo,** negative value of hippocampus volume PRS derived with Korean hippocampus GWAS; **-Amyg**, negative value of amygdala volume PRS derived with Korean amygdala GWAS; **-Entorh**, negative value of entorhinal thickness PRS derived with Korean entorhinal GWAS; **SBP**, SBP PRS derived with Japanese SBP GWAS; **BMI**, BMI PRS derived with Japanese BMI GWAS

Figure 4. Odds Ratio of each PRS with Alzheimer's Disease diagnosis status in Japanese test set



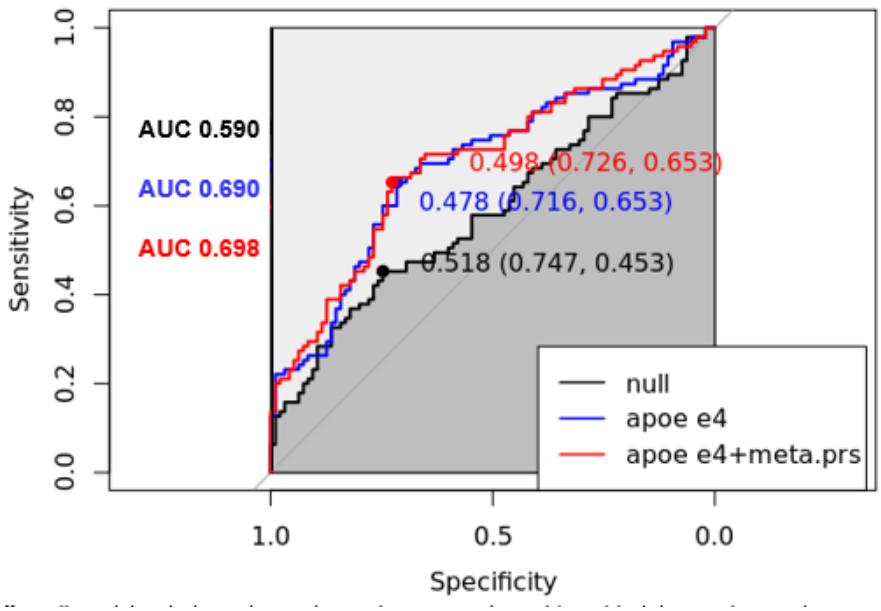
Shown are the Odds Ratio for Alzheimer's Disease diagnosis status in the Japanese test set. Models with meta-PRS including KOR.AD are shown in gray region, including EUR.AD are shown in blue region, and including both (KOR.AD and EUR.AD) are shown in red region. Odds Ratio of meta PRS (EUR.AD+(-Amyg)+(-Entorh)), whose AUC is the highest, is indicated by a red dashed line in the graph [OR=1.277, 95% CI 1.15-1.41]. Odds Ratio of KOR.AD [OR=1.008, 95% CI 0.91-1.12], and EUR.AD [OR=1.257, 95% CI 1.14-1.39] are indicated by black and blue dashed line in the graph, respectively. (Abbreviation) **KOR.AD**, AD PRS derived with Korean AD GWAS; **EUR.AD**, AD PRS derived with European AD GWAS; **Diag**, both AD PRSs used together (KOR.AD + EUR.AD); **(-Hippo)**, negative value of hippocampus volume PRS derived with Korean hippocampus GWAS; **(-Amyg)**, negative value of amygdala volume PRS derived with Korean amygdala GWAS; **(-Entorh)**, negative value of entorhinal thickness PRS derived with Korean entorhinal GWAS;

Figure 5. AUC of the model including each meta-PRS



Shown are the AUCs for Alzheimer's Disease diagnosis status classification in the Japanese test set. The APOE ε4 model included only APOE ε4 allele carrier status, gender, and 10 PCs, whose AUC was 0.690, indicated by a blue dashed line in the graph. The other models included each PRS, in addition. Models with meta-PRS including KOR.AD are shown in gray region, including EUR.AD are shown in blue region, and including both (KOR.AD and EUR.AD) are shown in red region. Best model was the meta PRS model of EUR.AD+(-Amyg)+(-Entorh), whose AUC was 0.698, indicated by a red dashed line. (Abbreviation) **KOR.AD**, AD PRS derived with Korean AD GWAS; **EUR.AD**, AD PRS derived with European AD GWAS; **Diag**, both AD PRSs used together (KOR.AD + EUR.AD); **(-Hippo)**, negative value of hippocampus volume PRS derived with Korean hippocampus GWAS; **(-Amyg)**, negative value of amygdala volume PRS derived with Korean amygdala GWAS; **(-Entorh)**, negative value of entorhinal thickness PRS derived with Korean entorhinal GWAS;

Figure 6. Comparison of AUC-ROC Curve



Shown are the AUCs and ROC curves of the null model, APOE ε4 model, and best meta PRS model for Alzheimer's Disease diagnosis status classification in the Japanese test set. (Abbreviation) **Null**, null model includes only gender, and 10 PCs, indicated by a black line in the graph; **APOE e4,** APOE ε4 model includes APOE ε4 carrier status, gender, and 10 PCs, indicated by a blue line in the graph; **APOE e4 + meta.prs,** APOE ε4 + meta.prs model includes the meta PRS(EUR.AD, negative value of Amygdala, and Entorhinal), which had the highest AUC among the meta PRSs, as well as gender and 10PCs, indicated by a red line in the graph;

Bibliography

1. Jiao, B., et al., Associations of risk genes with onset age and plasma biomarkers of Alzheimer's disease: a large case-control study in mainland China. Neuropsychopharmacology, 2022. 47(5): p. 1121-1127.
2. Harrison, J., et al., From Polygenic Scores to Precision Medicine in Alzheimer's Disease: A Systematic Review. Journal of Alzheimer's Disease, 2020. 74: p. 1-13.
3. Ridge, P.G., et al., Assessment of the genetic variance of late-onset Alzheimer's disease. Neurobiology of aging, 2016. 41: p. 200. e13-200. e20.
4. Escott-Price, V., et al., Polygenic risk score analysis of pathologically confirmed Alzheimer disease. Annals of neurology, 2017. 82(2): p. 311-314.
5. Li, J., et al., Polygenic risk for Alzheimer's disease influences precuneal volume in two independent general populations. Neurobiology of Aging, 2018. 64: p. 116-122.
6. Zhou, X., et al., Genetic and polygenic risk score analysis for Alzheimer's disease in the Chinese population. Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring, 2020. 12(1): p. e12074.
7. Park, J.-H., et al., Novel Alzheimer's disease risk variants identified based on whole-genome sequencing of APOE ε4 carriers. Translational psychiatry, 2021. 11(1): p. 1-12.
8. Miyashita, A., et al., SORL1 is genetically associated with late-onset Alzheimer's disease in Japanese, Koreans and Caucasians. PloS one, 2013. 8(4): p. e58618.
9. Hirano, A., et al., A genome-wide association study of late-onset Alzheimer's disease in a Japanese population. Psychiatric genetics, 2015. 25(4): p. 139-146.
10. Shigemizu, D., et al., Ethnic and trans-ethnic genome-wide association studies identify new loci influencing Japanese Alzheimer's disease risk. Translational psychiatry, 2021. 11(1): p. 1-10.
11. Zhou, X., et al., Identification of genetic risk factors in the Chinese population implicates a role of immune system in Alzheimer's disease pathogenesis. Proceedings of the National Academy of Sciences, 2018. 115(8): p. 1697-1706.
12. Jia, L., et al., Prediction of Alzheimer's disease using multi-variants from a Chinese genome-wide association study. Brain, 2021. 144(3): p. 924-937.
13. Abraham, G., et al., Genomic risk score offers predictive performance comparable to clinical risk factors for ischaemic stroke. Nature Communications, 2019. 10(1): p. 5819.
14. McKhann, G., et al., Clinical diagnosis of Alzheimer's disease. Neurology, 1984. 34(7): p. 939.
15. Bellenguez, C., et al., New insights into the genetic etiology of

Alzheimer's disease and related dementias. Nature Genetics, 2022. 54(4): p. 412-436.
16. Sakaue, S., et al., A cross-population atlas of genetic associations for 220 human phenotypes. Nature Genetics, 2021. 53(10): p. 1415-1424.
17. RexSoft, Rex: Excel-based statistical analysis software. 2018.
18. Moon, S., et al., The Korea Biobank Array: Design and Identification of Coding Variants Associated with Blood Biochemical Traits. Sci Rep, 2019. 9(1): p. 1382.
19. Seo, S., et al., SNP genotype calling and quality control for multi-batch-based studies. Genes & Genomics, 2019. 41(8): p. 927-939.
20. Purcell, S., et al., PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet, 2007. 81(3): p. 559-75.
21. Wigginton, J.E., D.J. Cutler, and G.R. Abecasis, A Note on Exact Tests of Hardy-Weinberg Equilibrium. The American Journal of Human Genetics, 2005. 76(5): p. 887-893.
22. Das, S., et al., Next-generation genotype imputation service and methods. Nature Genetics, 2016. 48(10): p. 1284-1287.
23. Yoo, S.-K., et al., NARD: whole-genome reference panel of 1779 Northeast Asians improves imputation accuracy of rare and low-frequency variants. Genome Medicine, 2019. 11(1): p. 64.
24. Park, J.Y., et al., A missense variant in SHARPIN mediates Alzheimer's disease-specific brain damages. Translational Psychiatry, 2021. 11(1): p. 590.
25. Kuznetsova, A., P.B. Brockhoff, and R.H.B. Christensen, lmerTest Package: Tests in Linear Mixed Effects Models. Journal of Statistical Software, 2017. 82(13): p. 1 - 26.
26. Watanabe, K., et al., Functional mapping and annotation of genetic associations with FUMA. Nature Communications, 2017. 8(1): p. 1826.
27. Privé, F., J. Arbel, and B.J. Vilhjálmsson, LDpred2: better, faster, stronger. Bioinformatics, 2020. 36(22-23): p. 5424-5431.
28. Friedman, J.H., T. Hastie, and R. Tibshirani, Regularization Paths for Generalized Linear Models via Coordinate Descent. Journal of Statistical Software, 2010. 33(1): p. 1 - 22.
29. Kang, S., et al., Potential novel genes for late-onset Alzheimer's disease in East-Asian descent identified by APOE-stratified genome-wide association study. Journal of Alzheimer's Disease, 2021. 82(4): p. 1451-1460.
30. Tang, X., et al., APOE Affects the Volume and Shape of the Amygdala and the Hippocampus in Mild Cognitive Impairment and Alzheimer's Disease: Age Matters. Journal of Alzheimer's Disease, 2015. 47: p. 645-660.
31. Miyashita, A., et al., Genetics of Alzheimer's disease: an East Asian perspective. Journal of Human Genetics, 2022: p. 1-10.
32. Wang, Y., et al., Genetic structure, divergence and admixture of Han Chinese, Japanese and Korean populations. Hereditas, 2018. 155(1): p. 19.

33. van Oostveen, W.M. and E.C.M. de Lange, Imaging Techniques in Alzheimer's Disease: A Review of Applications in Early Diagnosis and Longitudinal Monitoring. International Journal of Molecular Sciences, 2021. 22(4).
34. Leandrou, S., et al., Assessment of Alzheimer's Disease Based on Texture Analysis of the Entorhinal Cortex. Frontiers in Aging Neuroscience, 2020. 12.

국문초록

다유전자 위험 점수를 활용한 알츠하이머 질환 위험 예측 모델

임수진
서울대학교 보건대학원
보건학과 보건통계전공

연구 배경

치매의 대부분을 차지하는 알츠하이머병은 높은 유전율로 인해 관련 유전 요인을 밝히기 위한 전장유전체연구(GWAS)가 활발하게 진행되고 있다. 알츠하이머와 관련된 유전적 요인 중 가장 큰 연관성을 보이는 APOE를 제외하면, 각각의 유전 요인이 미치는 영향은 영향력은 작아, 알츠하이머에 대한 유전적 위험 예측에 있어 다유전자 위험 점수(PRS)를 활용한 접근이 적합하다. 따라서 알츠하이머 질환에 대한 GWAS 및 PRS 연구가 현재 활발히 진행 중이지만, 현재까지 보고된 연구들은 대부분 유럽인을 대상으로 하고 있다. 현재까지 진행된 동아시아인 기반 연구는 GWAS 7개이나, 이 중 APOE gene을 제외하면, 동일하게 보고된 유전적 변이가 없다. 이는 유럽 인종 연구에 비해 극히 작은 샘플 수로 통계적 검정력이 부족하기 때문으로 보인다. 정확한 유전적 위험 예측을 위한 PRS 식 개발에는 대규모 전장 유전체 연구가 필수적이기 때문에, 동아시아인 대상 PRS 연구는 현재까지 제한이 많다. 그러나 meta PRS와 같이, 새로운 통계적 접근법들을 활용해 부족한 샘플 수를 극복하기 위한 연구들이 활발하게 진행중에 있다.

연구 목적

본 연구에서는 meta PRS 방법을 동아시아인에서 알츠하이머병의 유전 위험 예측에 적용해보고 이를 기존의 전통적인 방법을 통해 산출된 PRS들과 비교

해 보고자 한다.

연구방법

본 연구에서는 한국인 치매 코호트에 등록된 60세 이상 한국인 4525명을 활용하여, 7가지 PRS 식을 개발하였다. 7가지 PRS 식은 다음과 같다; 한국인의 알츠하이머 진단에 대한 GWAS 분석 후 그 결과를 활용한 PRS(KOR.AD), 유럽인의 알츠하이머 진단에 대한 GWAS 요약 통계량을 활용한 PRS(EUR.AD), 한국인의 hippocampus, amygdala 부피와 entorhinal 두께에 대한 GWAS 분석 후 그 결과를 활용한 PRS(Hippocampus, Amygdala, Entorhinal), 그리고 일본인의 수축기 혈압 및 BMI에 대한 GWAS 요약 통계량을 활용한 PRS(SBP, BMI).

이후 각각의 PRS들의 21개 조합에 대한 21개 meta PRS는 일본인 치매 코호트를 둘로 나누어 개발하고 검증하였다(N=1,899). Meta PRS식의 산출은 알츠하이머 진단 여부에 대한 로지스틱 회귀식에 Ridge 페널티 모수를 추가로 적용한 후 10 fold cross validation을 통해 가장 좋은 예측 성능을 보이는 모수를 최적화했다. 이후 개발된 meta PRS식들은 전통적인 방법으로 개발된 알츠하이머 PRS인 KOR.AD 및 EUR.AD 독립된 테스트 셋에서 비교되었다.

연구 결과

최적의 성능을 보인 meta PRS는 EUR.AD, Amygdala, Entorhinal을 통합해 생성된 식이었다. 이는 알츠하이머 진단 상태와 유의한 양의 상관성을 보였으며(OR=1.277, 95% CI 1.15-1.41), 그 크기는 기존의 방식으로 산출된 KOR.AD(OR=1.008, 95% CI 0.91-1.12)와 EUR.AD(OR=1.257, 95% CI 1.14-1.39) 보다 크고 유의했으며, 이를 포함한 예측 모형의 예측 성능(AUC=0.698) 역시 KOR.AD(AUC=0.689)와 EUR.AD(AUC=0.695)를 포함한 예측 모형보다 우수했다. 21개의 meta PRS들 중 EUR.AD를 포함한 경우 그렇지 않은 경우보다 항상 알츠하이머 진단 상태에 대해 큰 연관성을 보였으며, 해당 meta PRS를 포함한 식의 예측성능이 항상 더 우수하였다. Meta

PRS들 중 KOR.AD가 포함된 경우, EUR.AD가 포함된 경우, 혹은 둘 다가 포함된 경우의 조건이 동일한 경우에는 Amygdala를 추가로 포함한 경우에서 항상 연관성 크고 예측성능이 우수하였다.

결론

본 연구는 동아시아인에서 알츠하이머에 대한 유전적 위험 점수를 산출한 두 번째 연구이며, meta PRS 방법을 적용함으로써, 비교적 작은 샘플 수를 이용해 알츠하이머 질환과 유의한 연관성을 보였다.

주요어 : 알츠하이머, 다유전자 위험 점수, 예측 모델, 전장유전체 분석
학　번 : 2021-22216