Novel polygenic risk score approach with transcriptome-based weighting for genetic risk prediction of late-onset Alzheimer's disease

Sang-Hyuk Jung[1,2], Kwangsik Nho, PhD[3], Li Shen[2], Dokyoon Kim, PhD[2],*, Hong-Hee Won, PhD[1], for the Alzheimer's Disease Neuroimaging Initiative

[1]Department of Digital Health, SAIHST, Sungkyunkwan University, Samsung Medical Center, Seoul, Republic of Korea; [2]Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, University of Pennsylvania, PA 19104, USA; [3]Department of Health Sciences and Technology, SAIHST, Sungkyunkwan University, Samsung Medical Center, Seoul, Republic of Korea.

Background

- Large-scale genome-wide association studies (GWAS) have identified more than 25 genomic regions related to Alzheimer's disease (AD).
- Recent studies showed polygenic risk score (PRS) could be used to identify individuals at high risk of AD. Despite this success, prediction and early intervention of AD still remain challenging.
- In this study, we suggest a novel PRS approach that is weighted by predicted tissue-specific gene expression levels.

Table 1. Demographics table

Status change (recruitment → onset)	Cognitive Normal	Late-onset Alzheimer's disease				
	-	CN → AD	EMCI[1)] → AD	LMCI[2)] → AD	AD	LOAD Total[3)]
N	239	**16** / 256 (6.25%)	**21** / 215 (9.77%)	**123** / 231 (53.25%)	**47**	**207**
Female (%)	119 (49.79)	7 (43.75)	6 (28.57)	47 (38.21)	17 (36.17)	77 (37.20) *
APOE ε4/ε4 (%)	134 (35.27)	8 (50.00)	13 (61.90)	80 (65.04)	33 (70.21)	134 (64.73) **
Age at recruitment (mean ± s.d.)	74.34 (± 5.57)	75.06 (± 3.86)	74.29 (± 5.90)	73.32 (± 7.16)	75.19 (± 9.31)	76.20 (± 7.81) *
Age at onset (AAO) (mean ± s.d.)		81.25 (± 3.86)	76.74 (± 6.56)	75.83 (± 7.58)		
Polygenic risk score (mean ± s.d.)	0.90 (± 0.29)	0.86 (± 0.24)	1.11 (± 0.30)	1.22 (± 0.33)		1.18 (± 0.33) **

Method

- We used whole-genome sequencing data of 446 European participants (207 AD cases and 239 cognitively normal controls) from the Alzheimer's disease Neuroimaging Initiative (ADNI), and performed quality control of genotyped data using PLINK (Table 1).

Table 2. Predictive performance of transcriptome-based weighting and conventional PRS in additive models

	ADNI cohort Late-onset Alzheimer's disease: 207 Cognitive normal: 239		Transcriptome-based weighting PRS (proposed)		ADNI cohort Late-onset Alzheimer's disease: 207 Cognitive normal: 239		Conventional PRS	
Model	**AUC**	**Pseudo R2**	**Beta (SE) (TW-PRS)**	***P*-value (TW-PRS)**	**AUC**	**Pseudo R2**	**Beta (SE) (PRS)**	***P*-value (PRS)**
1	0.6940	0.1924	.	.	0.6940	0.1924	.	.
2	0.7506	0.2429	0.4094 (0.0980)	2.94E-05	0.7472	0.2391	0.2439 (0.0607)	5.84E-05
3	0.6269	0.0539	.	.	0.6269	0.0539	.	.
4	0.7637	0.2599	.	.	0.7637	0.2599	.	.
5	**0.7829**	**0.3050**	0.4356 (0.1029)	**2.33E-05**	**0.7757**	**0.2975**	0.2446 (0.0630)	**0.0001**

Model 1: *APOE* ε4 status

Model 2: *APOE* ε4 status + PC1-3 + PRS (or TW-PRS)

Model 3: Sex + Age + Education

Model 4: Sex + Age + Education + *APOE* ε4 status + PC1-3

Model 5: Sex + Age + Education + *APOE* ε4 status + PC1-3 + PRS (or TW-PRS)

- We performed transcriptome-wide association studies (TWAS) in 13 brain regions by using MetaXcan (weights from GTEx V8) and GWAS summary statistics (IGAP stage 1 without ADNI samples) and integrated these 13 TWAS results using MultiXcan. To generate transcriptome-based weighting (TW)-PRS, expression weights in each gene were mapped to the SNP level, and these were applied as additional weights to the SNPs' beta coefficients in the GWAS summary statistics. Then, PRS was derived based on the weighted GWAS summary statistics using PRSice-2.
- Finally, we evaluated prediction performance of a set of different models that incorporated clinical features and weighted-PRS.

Result

- A total of 17,588 gene expression weights were obtained from 13 brain tissues. These gene expression weights covered about 220,000 SNPs. Among them, expression weights were applied to about 140,000 SNPs overlapped with GWAS summary statistics. An AD prediction using conventional PRS yielded a pseudo-R2 of 0.0462 (P=1E-04). Compared with the conventional PRS, the TBW-PRS improved the performance and statistical power with a pseudo-R2 of 0.0574 (P=1.74E-05). As shown in Table 2, a fully adjusted model achieved an AUC of 0.794 in which TW-PRS were significant even when APOE ε4 status and demographic information were adjusted (P=2.33E-05).

Conclusion

- This study proposes a novel TW-PRS approach that combines predicted tissue-specific transcriptomic weights and PRS. Expression weights of brain regions critical to AD progression enhanced the performance of conventional PRS. Our finding suggests that tissue-specific transcriptomic factors may be independent and complementary to conventional PRS and provide additional information for tissue-specific regulatory effects in AD.