



OPEN ACCESS

EDITED BY

Agustin Ibanez,
Latin American Brain Health Institute (BrainLat),
Chile

REVIEWED BY

Qiaojun Li,
Tianjin University of Commerce, China
Seyed Hani Hojjati,
Cornell University, United States
Sayantan Nag Chowdhury,
University of California, Davis, United States

*CORRESPONDENCE

Dokyoon Kim
✉ dokyoon.kim@penmedicine.upenn.edu

[†]These authors have contributed equally to this work and share first authorship

RECEIVED 23 August 2023

ACCEPTED 06 October 2023

PUBLISHED 26 October 2023

CITATION

Suh EH, Lee G, Jung S-H, Wen Z, Bao J, Nho K, Huang H, Davatzikos C, Saykin AJ, Thompson PM, Shen L and Kim D (2023) An interpretable Alzheimer's disease oligogenic risk score informed by neuroimaging biomarkers improves risk prediction and stratification. *Front. Aging Neurosci.* 15:1281748. doi: 10.3389/fnagi.2023.1281748

COPYRIGHT

© 2023 Suh, Lee, Jung, Wen, Bao, Nho, Huang, Davatzikos, Saykin, Thompson, Shen and Kim. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

An interpretable Alzheimer's disease oligogenic risk score informed by neuroimaging biomarkers improves risk prediction and stratification

Erica H. Suh^{1†}, Garam Lee^{1,2†}, Sang-Hyuk Jung¹, Zixuan Wen¹, Jingxuan Bao¹, Kwangsik Nho³, Heng Huang⁴, Christos Davatzikos⁵, Andrew J. Saykin³, Paul M. Thompson⁶, Li Shen^{1,7}, Dokyoon Kim^{1,7*} and for the Alzheimer's Disease Neuroimaging Initiative

¹Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States, ²Innovative Medical Technology Research Institute, Seoul National University Hospital, Seoul, Republic of Korea, ³Department of Radiology and Imaging Sciences, School of Medicine, Indiana University School of Medicine, Indianapolis, IN, United States, ⁴Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA, United States, ⁵Center for Biomedical Image Computing and Analytics, Department of Radiology, University of Pennsylvania, Philadelphia, PA, United States, ⁶Imaging Genetics Center, Mark and Mary Stevens Neuroimaging and Informatics Institute, Keck School of Medicine, University of Southern California, Los Angeles, CA, United States, ⁷Institute for Biomedical Informatics, University of Pennsylvania, Philadelphia, PA, United States

Introduction: Stratification of Alzheimer's disease (AD) patients into risk subgroups using Polygenic Risk Scores (PRS) presents novel opportunities for the development of clinical trials and disease-modifying therapies. However, the heterogeneous nature of AD continues to pose significant challenges for the clinical broadscale use of PRS. PRS remains unfit in demonstrating sufficient accuracy in risk prediction, particularly for individuals with mild cognitive impairment (MCI), and in allowing feasible interpretation of specific genes or SNPs contributing to disease risk. We propose adORS, a novel oligogenic risk score for AD, to better predict risk of disease by using an optimized list of relevant genetic risk factors.

Methods: Using whole genome sequencing data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) cohort ($n = 1,545$), we selected 20 genes that exhibited the strongest correlations with FDG-PET and AV45-PET, recognized neuroimaging biomarkers that detect functional brain changes in AD. This subset of genes was incorporated into adORS to assess, in comparison to PRS, the prediction accuracy of CN vs. AD classification and MCI conversion prediction, risk stratification of the ADNI cohort, and interpretability of the genetic information included in the scores.

Results: adORS improved AUC scores over PRS in both CN vs. AD classification and MCI conversion prediction. The oligogenic model also refined risk-based stratification, even without the assistance of APOE, thus reflecting the true prevalence rate of the ADNI cohort compared to PRS. Interpretation analysis shows that genes included in adORS, such as ATF6, EFCAB11, ING5, SIK3, and CD46, have been observed in similar neurodegenerative disorders and/or are supported by AD-related literature.

Discussion: Compared to conventional PRS, adORS may prove to be a more appropriate choice of differentiating patients into high or low genetic risk of AD

in clinical studies or settings. Additionally, the ability to interpret specific genetic information allows the focus to be shifted from general relative risk based on a given population to the information that adORS can provide for a single individual, thus permitting the possibility of personalized treatments for AD.

KEYWORDS

polygenic risk score, Alzheimer's disease, mild cognitive impairment, genetics, predictive markers

1. Introduction

Alzheimer's disease (AD) is the most common cause of dementia. It is an irreversible, progressive neurodegenerative disorder characterized by abnormal accumulation of amyloid plaques and neurofibrillary tangles in the brain, and disruption of memory, thinking, and behavior (Mullard, 2021). Because the root cause of AD remains unclear, current treatments only temporarily inhibit cognitive symptoms and fail to directly modify the course of the disease (Zetterberg and Bendlin, 2021). Limited progress of therapeutic development in clinical trials urges the need to consider alternative strategies for AD treatments, namely by investigating patients who are at risk of developing AD before disease onset.

Mild cognitive impairment (MCI)—a transitional stage between cognitively normal (CN) and early AD patients—may enable earlier detection of AD pathogenesis before clinical diagnosis (Alzheimer Association, 2021). Many investigators regard this phase as critical, for substantial neuronal damage has already occurred by the time the degenerative process of AD has fully developed. 15% of MCI individuals develop dementia after 2 years and 32% progress to AD within 5 years' follow-up, while others remain cognitively normal (Ward et al., 2013). The ability to predict progression from MCI to AD is critical for timely treatment to prevent or delay cognitive decline.

Polygenic risk scores (PRS) are a common approach for stratifying patients into levels of risk for AD by calculating the weighted sum of GWAS-significant SNP genotypes (Baker and Escott-Price, 2020; Sims et al., 2020; Leonenko et al., 2021; Gouveia et al., 2022). By considering the cumulative effects of many genetic variants together, one can better predict an individual's risk for a disease compared to considering single variants in isolation. Although AD's delayed symptom onset makes it difficult to manage, a deeper grasp of the underlying biology and early risk detection may pave the way for improved treatments for AD. This has, in part, motivated the application of PRS for AD (Marden et al., 2016; Desikan et al., 2017; Harrison et al., 2020; Leonenko et al., 2021; Jung et al., 2022). Previous studies have shown that PRS are able to quantify differences in genetic risk between individuals, enable risk-based stratification, and demonstrate strong associations with known markers of neurodegeneration (Desikan et al., 2017; Tan et al., 2019; Huq et al., 2021; Leonenko et al., 2021). However, despite longstanding evidence that AD risk is polygenic (Altmann et al., 2020), more recent studies have proposed that AD may in fact be an oligogenic disorder (Zhang et al., 2020). This leaves an open-ended discussion for how genetic risk for AD must be represented.

One primary limitation of PRS is that its calculation relies on genome-wide association studies (GWAS) of AD case-control datasets, which often use clinical diagnosis as phenotypes. However, AD clinical diagnosis partially relies on cognitive assessment outcomes, which may be influenced by factors unrelated to the disease

such as the patient's anxiety, fatigue, and general test-taking ability. Thus, it remains a challenge to identify genetic risk factors for AD solely using case-control-based GWASs. Additionally, GWAS studies examining MCI to AD progression indicate that AD susceptibility loci exhibit minimal effects. Current datasets with MCI individuals may lack the statistical power to identify significant SNPs, particularly with the inclusion of MCI subjects unlikely to develop AD or those progressing to other dementia types. Consequently, the observed effect sizes of true AD susceptibility genes in these MCI cohorts are further diminished compared to traditional AD case-control datasets, undermining the utility of PRS for individuals with MCI (Lacour et al., 2017; Chaudhury et al., 2019). Furthermore, PRS is not advocated for clinical use, as it not only is prone to bias, but also is unable to demonstrate sufficient specificity and sensitivity in risk prediction (Cuyvers and Slegers, 2016).

Conversely, neuroimaging data, also commonly used for AD diagnosis, e.g., magnetic resonance imaging (MRI), 2-Deoxy-2-[¹⁸F] fluorodeoxyglucose (FDG)-PET, and [¹⁸F] Flortetapir (AV45)-PET, are a highly reproducible and more objective data source (Sperling et al., 2011; Suppiah et al., 2019). FDG-PET tracks glucose metabolism and is particularly useful for identifying areas of the brain distinctly impacted by AD, although its accuracy may decline in older patients (Foster et al., 2007; Sperling et al., 2011). AV45-PET specifically visualizes amyloid-beta plaques, a key hallmark of AD, offering a significant advancement over previous methods that could only confirm these plaques post-mortem (Suppiah et al., 2019). Compared to other diagnostic measures like cerebrospinal fluid (CSF) tests and cognitive assessments, these PET imaging techniques are offer a real-time view of critical pathological features in the brain.

Recent studies have demonstrated that neuroimaging biomarkers can effectively predict AD progression. Follow-up data of AD patients exhibited reduced FDG uptake in the frontal, parietal and lateral temporal lobes, indicating strong biomarkers for MCI conversion prediction (Ossenkoppele et al., 2012; Caminiti et al., 2018). Huang et al. (2017) used longitudinal MRI data to classify MCI subjects as either MCI converters (MCI-c) or MCI non-converters (MCI-nc), achieving 79.4% accuracy. Integrating MRI with FDG-PET data increased prediction performance to 86.4% (Lu et al., 2018; Lee et al., 2019). These findings demonstrate that neuroimaging endophenotypes have the potential to reflect genotypes correlated with brain structure and function and may also help characterize the mechanisms of such risk alleles or genes associated with AD risk. Identifying SNPs or genes that affect or are correlated with changes in neuroimaging traits can further our knowledge of the genetic underpinnings of AD risk and progression.

As such, we generated adORS, a novel oligogenic risk score for Alzheimer's disease, to robustly predict high or low risk of disease and the likelihood of MCI-to-AD conversion. This risk score leverages neuroimaging biomarkers FDG-PET and AV45-PET (Shen and

Thompson, 2020) to extract a smaller, more informative set of AD-associated genes to ultimately improve risk prediction performance for two separate tasks: CN vs. AD classification (CNAD) and MCI conversion prediction (MCI-CP). We compared prediction performance between adORS and conventional PRS and interpreted the genetic contributions within each score.

2. Materials and methods

2.1. ADNI and ADSP-FUS1-ADNI-WGS-2 cohorts

Alzheimer's Disease Neuroimaging Initiative (ADNI; ADNI-WGS-1, 2020) is a multisite longitudinal study that tracks the progression of AD in the human brain with clinical, imaging, genetic and biospecimen biomarkers through the process of normal aging, early mild cognitive impairment, and late mild cognitive impairment to dementia or AD. The overall goal of ADNI is to validate biomarkers for use in Alzheimer's disease clinical treatment trials. ADSP-FUS1-ADNI-WGS-2 indicates an additional set of ADNI participants that underwent whole genome sequencing (WGS) as a collaboration between ADNI and ADSP Follow-Up Study (FUS; ADSP FUS1 WGS, 2021). We focus on the use of WGS data (see section 2.2) and baseline FDG-PET and AV45-PET biomarkers (see section 2.3), which are inherently cross-sectional. Hence, this a cross-sectional study. We collected data from 1,545 individuals across all modalities.

In the ADNI study, each participant receives one of four diagnoses during each visit: CN, early MCI, late MCI, or AD. Follow up data post baseline occurs every 6 months and average about 5 years in duration. To transform the longitudinal diagnoses into a cross-sectional format, we assigned each individual with a single, cumulative label based on their diagnosis history from all available visits. The labels are as follows for the two separate tasks:

1. CNAD task: AD cases (AD) and cognitively normal controls (CN).
2. MCI-CP task: MCI converters (MCI-c) and MCI non-converters (MCI-nc).

We generated these definitions for each label:

- MCI-c: Diagnosed with early or late MCI at baseline and later progresses to AD, maintaining this until the last visit.
- MCI-nc: Diagnosed with early or late MCI at baseline and remains with this diagnosis through the final visit.
- AD: Either consistently diagnosed with AD or categorized as MCI-c.
- CN: Remains cognitively normal from the first to the last visit.

Using these definitions, we acquired 550 AD cases, 364 CN controls, 273 MCI converters, and 517 MCI non-converters (Table 1).

2.2. WGS genotyping and quality control

A total of 1,566 samples were genotyped from ADNI and ADNI-WGS-2 with ADSP Follow-Up Study. WGS genotyping was performed using HiSeq2000; Read length (bp) 100; genome assembly GRCh38 (hg38). Quality Control {QC; SNV concordance check, sex mismatch,

contamination check, and relatedness [(Pihat >0.4)] were performed primarily using the SNP/Indel Variant Calling Pipeline (VCPA) v1.0 developed and maintained by National Institute on Aging Genetics of Alzheimer's Disease Data Storage Site (NIAGADS; Leung et al., 2019). All QC quality checks were implemented by the Genome Center for Alzheimer's Disease (GCAD) Data Production Team. And variants were filtered out with low-quality, multi-allelic, and monomorphic SNVs using filters of GQ < 20 and DP < 10 and a missing rate > 20% and a $p < 1.0 \times 10^{-6}$ for the Hardy-Weinberg equilibrium test. After QC, 15,456,635 variants and 1,545 samples remained available for performing PRS and adORS in the ADNI and ADSP-FUS1-ADNI-WGS-2 datasets.

2.3. FDG-PET and AV45-PET

Every FDG and AV45 PET scan is reviewed for protocol compliance by the ADNI PET QC team. PET measures are reported as standard uptake value ratios (SUVRs), which is a quantification of the radiotracer concentration within specific region of interests (ROI) in the body (ADNI PET analysis method, 2023). The average AV45 SUVR was computed from regions including the frontal, anterior cingulate, precuneus, and parietal cortex relative to the cerebellum. For FDG-PET, the average SUVR was based on the uptake in the angular, temporal, and posterior cingulate regions, areas known to exhibit altered metabolic activity in Alzheimer's disease. Such measures help elucidate the metabolic changes associated with the disease's progression.

2.4. PRS calculation

To calculate PRS, we used summary statistics from a large-scale GWAS study done for AD on the International Genomics of Alzheimer's Project (IGAP) consortium to generate genetic scores for ADNI participants. The IGAP GWAS was re-performed excluding 441 ADNI participants (55,931 participants remained across the Alzheimer's Disease Genetic Consortium (ADGC), the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium, the European Alzheimer's Disease Initiative (EADI), and the Genetic and Environmental Risk in Alzheimer's Disease (GERAD) Consortium). We generated PRS with the PLINK genetic data analysis toolset using clumping and thresholding (C+T), a common method which removes redundantly correlated variants and preferentially retains those most associated with the phenotype of interest. This method was chosen based on the suggested method of PRS calculation and threshold cutoffs performed in Leonenko et al. (2021). In our calculations, SNPs in the APOE region were first removed (chr19:44.0–46.0 Mb, GRCh38), then clumped by excluding nearby variants within a 1,000-kb window with $r^2 > 0.1$, and finally filtered using the threshold $pT \leq 1e-5$. We also computed additional PRSs using $pT \leq 5e-8, 0.1, 0.5$ (Supplementary Table 1). Given that SNPs in the APOE region were removed from the PRS calculation, we modeled APOE separately as the weighted sum of alleles $\epsilon 2$ and $\epsilon 4$ using effect sizes found in the IGAP summary statistics [$\beta(\epsilon 2) = -0.47, \beta(\epsilon 4) = 1.12$].

2.5. adORS calculation

Our proposed method is comprised of three steps for adORS calculation: gene-level aggregation, gene selection, and model

TABLE 1 ADNI and ADSP-FUS1-ADNI-WGS-2 data demographics.

	Cohort (top: control, bottom: case)	Cohort definition	N samples	Mean age (SD)	% Males/Females	APOE distribution (ε2ε2/ε2ε3/ε2ε4/ε3ε3/ε3ε4/ε4ε4)	Mean cognitive test scores (SD)				
							ADAS13	CDRSB	FAQ	MMSE	RAVLT
CN vs. AD classification	Cognitive normal	CN diagnosis from first to last visit	364	74.69 (7.47)	50.82/49.18	2/46/2/214/89/10	9.34 (4.34)	0.03 (0.13)	0.17 (0.63)	29.06 (1.15)	5.81 (2.28)
	Alzheimer's disease	CN or AD diagnosis at first visit, AD at last visit	550	74.32 (5.74)	58.91/41.09	1/17/10/174/252/96	25.55 (8.56)	3.19 (1.87)	9.48 (7.26)	25.03 (2.74)	2.43 (2.08)
MCI conversion prediction	MCI non-converters	EMCI or LMCI diagnosis from first to last visit	517	72.48 (7.81)	58.41/41.59	1/37/14/263/159/42	14.07 (5.89)	1.28 (0.73)	1.92 (3.06)	28.01 (1.69)	4.72 (2.54)
	MCI converters	EMCI or LMCI diagnosis at first visit, AD at last visit	273	73.84 (7.01)	62.27/37.73	0/8/6/92/124/43	20.91 (6.20)	1.90 (0.98)	5.38 (4.85)	26.97 (1.77)	2.99 (2.22)

Phenotypic information available in ADNI for each classification problem: cohort cases and controls, criteria used to define cases and controls, number of samples, mean age and standard deviations, APOE distribution, and mean cognitive scores and standard deviations (see Section 2.7 for details on cognitive scores). CN, Cognitively normal; AD, Alzheimer's disease; MCI, Mild cognitive impairment; EMCI, Early mild cognitive impairment; LMCI, Late mild cognitive impairment.

training. First, SNPs of each patient were aggregated into their corresponding gene burdens. It is denoted in Equations (1, 2) and where v_{ij} represents j -th gene burden of patient i . s_{ik} in Equation (1) is the transformation of k -th variant:

$$s_{ik} = \begin{cases} 2 - x_{ik}, & \text{if negative association with AD} \\ x_{ik}, & \text{otherwise} \end{cases} \quad (1)$$

$$v_{ij} = \sum_k^{g_j} s_{ik} \quad (2)$$

Gene burdens were simply computed by summing each SNP or complement of SNP in a given gene depending on the direction of each variant's effect on Alzheimer's disease. For example, a variant is summed after complement operation if the variant is negatively associated with the AD while regular variant is added otherwise. This prevents variants in opposite direction from canceling out each variants' effects. Genes were defined using refGene database and the ANNOVAR tool. Second, gene selection is performed to condense the number of genes included in calculating adORS. We measured Pearson correlation coefficients with each neuroimaging biomarker, FDG-PET and AV45-PET, and then, extracted the top 20 genes exhibiting the highest correlation (Supplementary Figure 1; Supplementary Table 2). Since FDG-PET and AV45-PET have been known as promising biomarkers to detect functional brain changes in AD, we used them as intermediate phenotypes to discover AD-relevant genes (Foster et al., 2007; Sperling et al., 2011; Suppiah et al., 2019). After accounting for overlapping genes from each set of 20 genes, the total of 37 genes were used to calculate adORS. Finally, the risk score for each patient i was estimated based on the final condensed gene set G :

$$adORS_i = \sum_j^G \beta_j v_{ij} \quad (3)$$

β_j in Equation (3) is the weight associated with each gene burden v_{ij} , calculated by a logistic regression (LR) model trained on the task of interest. Given that we focus on two problems, CNAD and MCI-CP, a separate LR model was trained for each task. As mentioned in section 2.1, only CN and AD subjects were used to train the CNAD LR model, and correspondingly, only MCI-nc and MCI-c subjects were used to train the MCI-CP model. Each LR model was implemented from the sklearn package in Python and adjusted for age and gender. They were developed and assessed using an 80:20 train test split and 10-fold cross-validation, with each fold containing the same ratio of positive and negative samples from the ADNI cohort. Prediction accuracy of the PRS and adORS models was measured using the Area Under Receiver Operating Characteristic (AUROC) and Area Under the Precision-Recall Curve (AURPC).

2.6. Gene contribution

Gene contribution of adORS was calculated using Shapley Additive exPlanations (SHAP; Lundberg and Lee, 2017). SHAP is a model-agnostic interpretation method frequently used in deep neural network and tree-based ensemble models. The strategy of SHAP is to

TABLE 2 Descriptions of PRS and adORS models presented in the manuscript.

Model name	Model description
APOE.only	Weighted sum of alleles $\epsilon 2$ and $\epsilon 4$ using effect sizes found in the IGAP summary statistics [$\beta(\epsilon 2) = -0.47$, $\beta(\epsilon 4) = 1.12$]
PRS	PRS including SNPs with $p \leq 1e-5$ and excluding SNPs in the APOE region (chr19:44.0–46.0 Mb, GRCh38)
PRS.with.APOE	PRS including SNPs with $p \leq 1e-5$ and APOE($\epsilon 2 + \epsilon 4$), where APOE effects were weighted with effect sizes [$\beta(\epsilon 2) = -0.47$ and $\beta(\epsilon 4) = 1.12$] from the IGAP summary statistics excluding the ADNI cohort
adORS	Risk score including the top 20 gene burdens with highest correlations to AV45-PET and FDG-PET and excluding the APOE region (chr19:44.0–46.0 Mb, GRCh38)
adORS.with.APOE	Risk score including the top 20 gene burdens with highest correlations to AV45-PET and FDG-PET and APOE($\epsilon 2 + \epsilon 4$), where APOE effects were weighted with effect sizes [$\beta(\epsilon 2) = -0.47$ and $\beta(\epsilon 4) = 1.12$] from the IGAP summary statistics excluding the ADNI cohort

TABLE 3 adORS vs. PRS performance comparisons for CN vs. AD classification and MCI conversion prediction.

Model	AUROC (SD)	AUPRC (SD)
(a) CN vs. AD classification		
APOE.only _{CNAD}	0.707 (0.04)	0.746 (0.03)
PRS _{CNAD}	0.602 (0.03)	0.694 (0.03)
PRS.with.APOE _{CNAD}	0.735 (0.04)	0.814 (0.03)
adORS _{CNAD}	0.634 (0.05)	0.717 (0.04)
adORS.with.APOE _{CNAD}	0.750 (0.03)	0.815 (0.03)
(b) MCI conversion prediction		
APOE.only _{MCI-CP}	0.594 (0.05)	0.417 (0.04)
PRS _{MCI-CP}	0.532 (0.04)	0.405 (0.03)
PRS.with.APOE _{MCI-CP}	0.643 (0.04)	0.477 (0.04)
adORS _{MCI-CP}	0.610 (0.04)	0.464 (0.05)
adORS.with.APOE _{MCI-CP}	0.637 (0.04)	0.487 (0.05)

All models were calculated on the ADNI cohort (550 AD cases, 364 CN controls, 273 MCI converters, and 517 MCI non-converters). See the section 2 for details on adORS and PRS calculation. For each task, five models were considered: APOE.only which is the weighted sum of APOE alleles $\epsilon 2$ and $\epsilon 4$ using summary statistics; PRS where the APOE region was excluded (chr19:44.0–46.0 Mb, GRCh38); PRS.with.APOE where APOE effects were weighted with effect sizes as in the summary statistics; adORS which includes the top 20 gene burdens with highest correlations to AV45-PET and FDG-PET; and adORS.with.APOE which also includes weighted APOE effects. Prediction performance was evaluated using AUROC and AUPRC with the standard deviation of 10-fold cross validation.

measure features’ relative effect contributing to model’s output while considering not only isolated effect of a single feature but also interactions between features. We applied SHAP package to trained adORS model to generate feature importance, namely SHAP value. To calculate gene contribution of PRS, we transformed variant-level into gene-level contribution. Since there is no gene-level summary statistics in PRS, we first calculated variant-level contribution using summary statistics multiplied by variant ($\beta_k x_{ik}$), and then the measures of variants’ contribution in a given gene were summed. Instead of investigating coefficient (β_k), we defined $\beta_k x_{ik}$ as a contribution measure of k -th variants because it not only reflects actual contribution to the level of risk score, but also enables to compute gene-level feature importance.

2.7. Phenotype association

Five scores (ADAS13, CDRSB, RAVLT, MMSE, and FAP) assessing the abilities of orientation to place and time, language,

reasoning, and the subjects’ daily living activities were downloaded from Quantitative Templates for the Progression of Alzheimer’s disease (QT-PAD) challenge (Portland Institute for Computational Science, 2022). We preserved the baseline visit for each subject (Table 1). We matched the subjects in our adORS analysis and QT-PAD baseline visit. Subjects without any of the five cognitive scores were excluded. The training and testing subjects are matched with our adORS analysis. We calculated Pearson correlations between adORS/PRS and five cognitive function assessment scores.

3. Results

3.1. adORS vs. PRS model performance comparison

To evaluate adORS’s ability to predict AD status and progression, we compared classification performance of adORS to PRS using the following evaluation metrics: Area Under Receiver Operating Characteristic (AUROC) and Area Under the Precision-Recall Curve (AUPRC). Performance was compared on two tasks: CN vs. AD classification and MCI conversion prediction. Table 2 details the description of each model. The PRS and adORS models were computed excluding SNPs in the APOE region (see section 2). Two additional models were generated by adding APOE as a separate independent variable to the feature set. APOE effects were modeled by summing the genotyped APOE isoforms $\epsilon 2$ and $\epsilon 4$ weighted with GWAS effect sizes from the IGAP summary statistics (Kunkle et al., 2019).

As presented in Table 3, adORS_{CNAD} (AUROC = 0.634, AUPRC = 0.717) achieves higher accuracy than PRS_{CNAD} (AUROC = 0.602, AUPRC = 0.694). This suggests that variants used to calculate adORS_{CNAD} hold higher predictive power than those used in PRS. When adding the APOE effect to both models, the performances of PRS.with.APOE_{CNAD} and adORS.with.APOE_{CNAD} increase to AUROC = 0.735 and AUROC = 0.750, respectively. The sharp increase in performance from PRS_{CNAD} (AUROC = 0.602) to PRS.with.APOE_{CNAD} (AUROC = 0.735) suggests that the APOE loci may be contributing most to prediction, causing other variants to have low effects. While a similar increase in performance was observed from adORS_{CNAD} (AUROC = 0.634) to adORS.with.APOE_{CNAD} (AUROC = 0.750), the higher performance values from the adORS models overall, and in comparison to APOE.only (AUROC = 0.707), demonstrate that additional genes are contributing to AD prediction more than they may be in the PRS models (Table 3). We analyze this phenomenon further in the Gene contribution section.

Similar patterns were observed in the MCI-CP task. Although the performance of PRS_{MCI-CP} was much lower (AUROC=0.532, AUPRC=0.405) than in the case of PRS_{CNAD}, it comparably increased when the *APOE* region was included (AUROC=0.643, AUPRC=0.477). In contrast, only a small increase in performance was observed from adORS_{MCI-CP} (AUROC=0.610, AUPRC=0.464) to adORS.with.*APOE*_{MCI-CP} (AUROC=0.637, AUPRC=0.487). This smaller increase in performance in the adORS models further emphasizes the idea that adORS relies less so on the *APOE* loci for the prediction task.

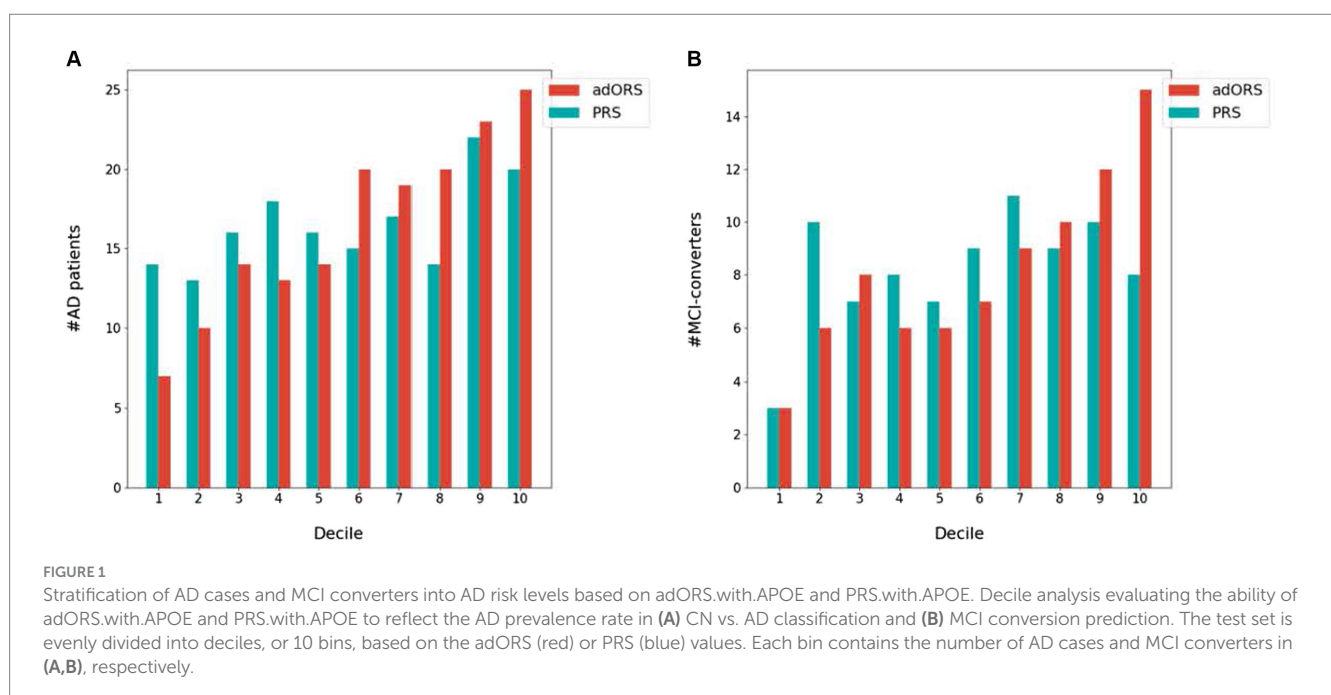
3.2. Risk-based stratification

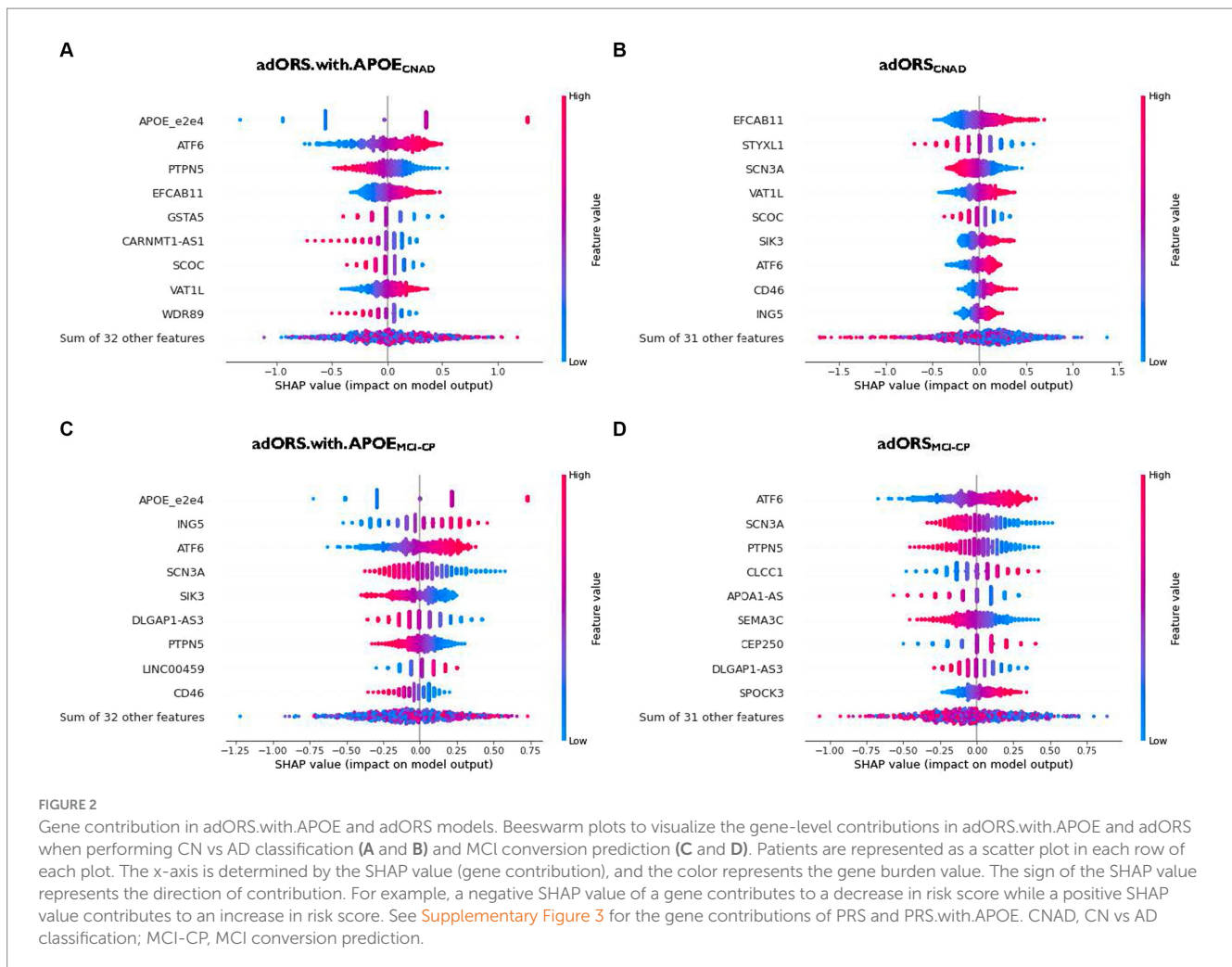
We next investigated each risk score's ability to stratify the cohort into (Mullard, 2021) cases vs. controls and (Zetterberg and Bendlin, 2021) bins of increasing AD risk levels. We performed a decile analysis to evaluate how well each score truly represents the AD prevalence rate. Figure 1 presents the test set divided into 10 equally sized bins based on risk score values and the number of actual cases within each bin. In both the CNAD (Figure 1A) and MCI-CP (Figure 1B) tasks, adORS demonstrates a staircase effect from the first to last decile, indicating that the score is well-calibrated to risk level compared to PRS. As the adORS model assigns higher risk scores to patients, the number of cases in higher risk groups increases accordingly. However, the decile of risk scores binned by PRS appears less as a staircase effect and more uniform across all bins. This indicates that PRS hardly reflects the true prevalence rate within ADNI cohort. Similar trends are seen for the adORS and PRS models which do not include *APOE* (Supplementary Figure 2). Overall, adORS's ability to improve stratification of patients into appropriate risk groups demonstrates that adORS is more well-calibrated to AD risk than PRS.

3.3. Gene contribution

The top contributors of adORS.with.*APOE*_{CNAD} and adORS_{CNAD} are *APOE* and *EFCAB11*, respectively, given that both genes demonstrate the highest magnitude of the SHAP value (Figures 2A,B; Supplementary Figures 3A,B). There is no overlap between the subsequent top five contributors across both models, excluding *EFCAB11*. This lack of overlap implies that the top genes found in adORS_{CNAD} may have associations with *APOE*, given that they are assumedly replacing *APOE*'s role. *APOE* also remains the top contributor for adORS.with.*APOE*_{MCI-CP} (Figure 2C; Supplementary Figure 3C). In adORS_{MCI-CP}, *ATF6* is the top contributor, with *SCN3A* being the second-most contributor (Figure 2D; Supplementary Figure 3D). *ATF6* and *SCN3A* are also part of the top five contributors in adORS.with.*APOE*_{MCI-CP}. There is very little overlap in gene contributions between the adORS and PRS models in both tasks (Figure 2; Supplementary Figures 3E–H), indicating that a different set of genes were considered to calculate PRS versus adORS.

As we reported in the *adORS* vs. *PRS* model performance comparison section, performance metrics suggested that PRS relies on *APOE* more so than in adORS, causing remainder variants to have low effects. As such, interpretation of genes—beyond *APOE*—that add to AD risk becomes more difficult for the PRS models. We further explored this notion using our gene contribution analysis, as shown in Supplementary Figure 3. The contribution of *APOE* in adORS.with.*APOE*_{CNAD} vs. PRS.with.*APOE*_{CNAD} is 19.6 and 20.9%, respectively (Supplementary Figures 3A,E). While these values are comparable, the distribution of contributions across the remainder of the genes is more dispersed in adORS.with.*APOE*_{CNAD}. Most of the genes excluding *APOE* contribute about 2.3–7.2% to the prediction task (Supplementary Figure 3A), whereas in PRS.with.*APOE*_{CNAD}, contributions range from 1.7 to 4.7%, with more than half of the genes contributing less than 1.7%





([Supplementary Figure 3E](#)). For MCI-CP, the difference is even more evident—*APOE* contributes 13.0% in adORS.with.APOE_{MCI-CP} and 62.5% in PRS.with.APOE_{MCI-CP} ([Supplementary Figures 3C,G](#)). The others contribute 2.2–6.9%, while in PRS, the top gene besides *APOE* contributes 2.3% at most, with nearly the rest contributing less than 1%.

We also investigated the direction of effect between the gene burdens and risk scores. In CN vs. AD classification, genes including *APOE*, *ATF6*, *EFCAB11*, *VAT1L*, *SIK3*, *CD46*, and *ING5* demonstrate a consistent, positive effect on the risk score ([Figures 2A,B](#)). Their SHAP values increase as their original values, or gene burdens, also increase. That is, higher gene burden values elicit higher risk for AD, and vice versa. In MCI conversion prediction, *ING5*, *ATF6*, *LINC00459*, *CLCC1*, *CEP250*, and *SPOCK3* are observed to be positive contributors, while *SCN3A*, *SIK3*, *DLGAP1-AS3*, *PTPN5*, *CD46*, *APOA1-AS*, and *SEMA3C* are negative contributors ([Figures 2C,D](#)).

Interestingly, *SIK3* and *CD46* display an inconsistent direction of effect on the risk score across each task. *SIK3* in adORS.with.APOE_{MCI-CP} for example, exhibits a negative effect on the risk score, yet the effect is reversed in adORS_{CNAD} ([Figures 2B,C](#)). The same trends are observed for *CD46*. This reversal of effect from MCI-CP to CNAD suggests that both genes may play an important role in the progression of AD.

3.4. Phenotype association

We performed an association analysis on PRS and adORS with major AD clinical assessments of cognitive functions (ADAS13, CDRSB, RAVLT, MMSE, and FAQ; [Portland Institute for Computational Science, 2022](#)) widely used as evidence for AD diagnosis ([Albert et al., 2011](#)). We assessed the Pearson correlations between the estimated risk scores and these cognitive assessment scores ([Figure 3](#)). Both PRS_{CNAD} and adORS_{CNAD} appropriately showed a positive correlation with ADAS13/CDRSB/FAQ ([Figure 3B](#)). We also observed a negative correlation between the estimated risk scores and MMSE/RAVLT, which is as expected. Compared to PRS_{CNAD}, adORS_{CNAD} had almost 2-fold higher absolute correlations. When *APOE* was included, adORS.with.APOE_{CNAD} still remained more strongly correlated than did PRS.with.APOE_{CNAD}, yet the trends of PRS.with.APOE_{CNAD}'s correlation values improve vastly in comparison to PRS_{CNAD} ([Figure 3A](#)). This evidence further supports the assumption that PRS heavily weights the *APOE* loci over other variants to sufficiently predict risk.

In general, MCI conversion prediction is more difficult than CN vs. AD classification. Correlations varied between adORS.with.APOE_{MCI-CP} and PRS.with.APOE_{MCI-CP} with adORS.with.APOE_{MCI-CP} performing better for MMSE while PRS.with.APOE_{MCI-CP} performed better for ADAS13/RAVLT ([Figure 3C](#)). This is consistent with the

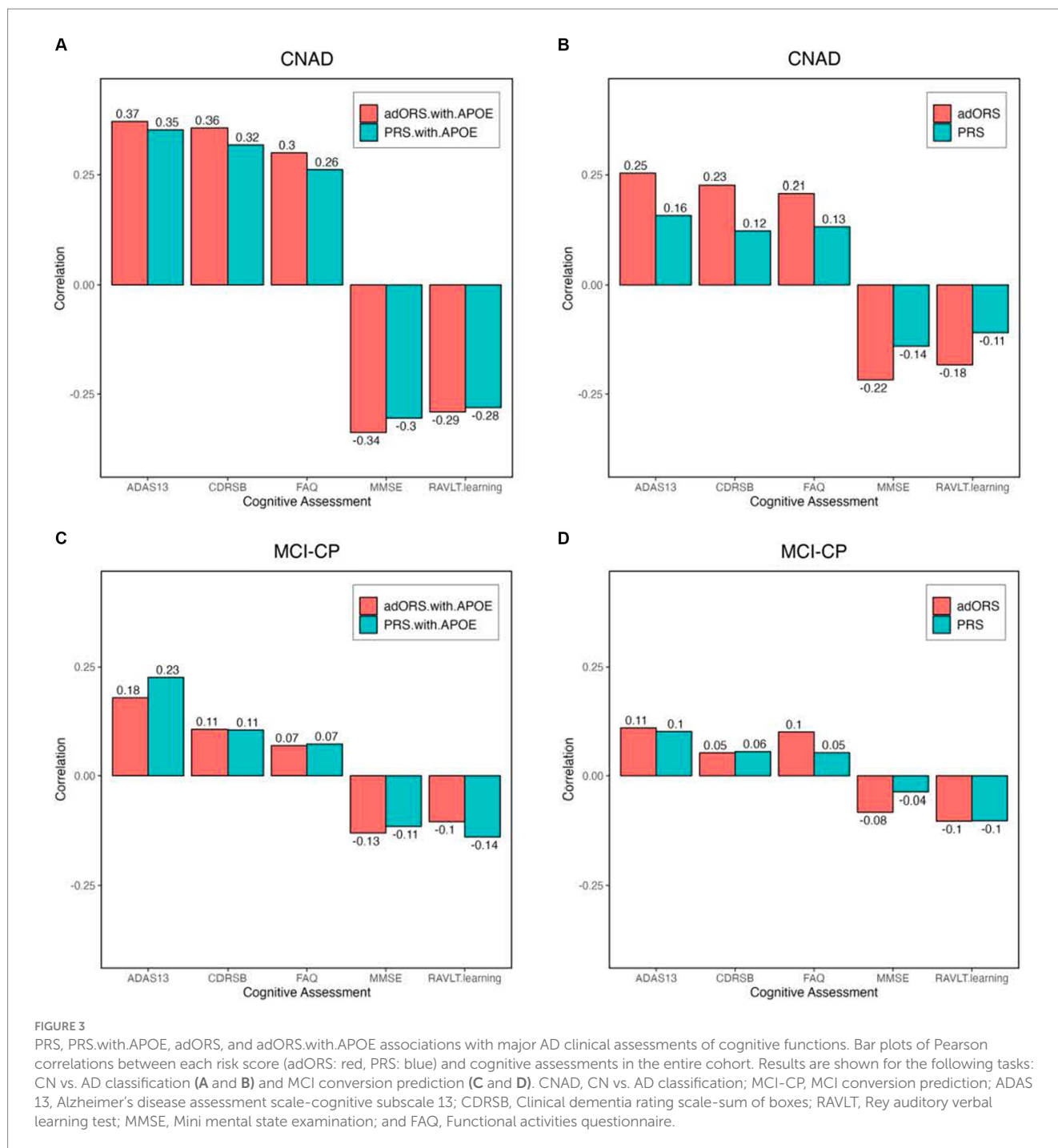


FIGURE 3 PRS, PRS.with.APOE, adORS, and adORS.with.APOE associations with major AD clinical assessments of cognitive functions. Bar plots of Pearson correlations between each risk score (adORS: red, PRS: blue) and cognitive assessments in the entire cohort. Results are shown for the following tasks: CN vs. AD classification (A and B) and MCI conversion prediction (C and D). CNAD, CN vs. AD classification; MCI-CP, MCI conversion prediction; ADAS 13, Alzheimer's disease assessment scale-cognitive subscale 13; CDRSB, Clinical dementia rating scale-sum of boxes; RAVLT, Rey auditory verbal learning test; MMSE, Mini mental state examination; and FAQ, Functional activities questionnaire.

model performances shown in Table 3, where PRS.with.APOE_{MCI-CP} resulted in a higher AUROC than adORS.with.APOE_{MCI-CP}. Without APOE, adORS_{MCI-CP} improved in comparison to PRS_{MCI-CP} with comparable yet better correlations in all cognitive categories except for CDRSB (Figure 3D).

4. Discussion

Given the lack of successful clinical trials for Alzheimer's disease to date, the field has shifted some of its focus to genetic risk stratification using PRS in hopes to improve mechanistic insights of

the disease (Sierksma et al., 2020). However, whether AD is a polygenic or oligogenic disorder remains unclear, and many suspect PRS's ability to accurately identify high and low risk individuals, as well as effectively predict AD progression. In the current study, we generated adORS, an AD-specific oligogenic risk score that challenges the limitations of PRS in terms of prediction, stratification, and interpretability. Whereas PRS is derived from the weighted sum of GWAS-significant SNPs, adORS utilizes genes strongly associated with the neuroimaging biomarkers FDG-PET and AV45-PET to predict both AD and MCI conversion risk. The introduced score outperforms traditional PRS by providing enhanced prediction of AD risk. Additionally, adORS better stratifies individuals into increasing

risk levels based on their genetic predispositions, reflecting the true prevalence rate of the cohort compared to PRS. Lastly, adORS delves into gene-level interpretations of the risk factors, paving the way for personalized therapeutic strategies for individuals with AD.

We conclude that the empirical data assessing AD risk are consistent with an oligogenic architecture of the disease. To represent oligogenicity of the disease, we implemented a gene-level burden analysis which collapses variants into genes. Because SNP data are extremely sparse and sample size is limited, we attempted to enlarge the gene-level effects that are relevant to AD by significantly reducing the feature size from 1.7 million SNPs to 23,000 genes. In other words, this approach allowed a denser input which permits the prediction model to train with increased stability. We expected several main effects from this architecture: enrichment of the AD-related signal, reduction of the degree of freedom, and gene-level interpretability. We anticipate that these outcomes may help establish the prospective clinical utility of adORS which PRS has been lacking thus far.

Compared to PRS, we found that the oligogenic model better represented the genetic heterogeneity of the disease by improving risk-based stratification for both CN vs. AD classification and MCI conversion prediction (Figure 1). When APOE was included in both models, the staircase effect portrayed in Figure 1 demonstrated that adORS is well-calibrated to AD risk level, whereas the decile of risk scores binned by PRS appears less as a staircase effect and more uniform across all bins. In other words, as adORS assigned higher risk scores to patients, the number of cases in higher risk groups increases accordingly. By refining stratification, adORS may prove to be a more appropriate choice of differentiating patients into high or low genetic risk of AD in clinical studies or settings. We were able to further support adORS's potential clinical utility via our association analysis with common cognitive assessments used to clinically evaluate risk for AD (Figure 3).

However, while APOE is the most prominent genetic risk factor for AD, not all AD patients carry the APOE risk allele (Guerreiro et al., 2012). As such, we deemed it essential to conduct analyses both inclusive and exclusive of APOE effects. The conventional PRS model has been evidenced to heavily depend on APOE and nongenetic factors, such as age and sex, for effective risk assessment, which compromises its clinical relevance (Riedel et al., 2016; Harrison et al., 2020; Leonenko et al., 2021). When APOE effects were excluded from the adORS and PRS models, adORS still improved the stratification of patients from lowest to highest risk group, while PRS continued to disperse patients more uniformly across all risk levels (Supplementary Figure 2). Moreover, the overwhelming contribution of APOE in PRS made it challenging to gauge the significance of individual SNPs or genes, especially in predicting MCI conversion (Supplementary Figures 3E,G). This disproportionate contribution by APOE obscures the roles of other genetic components, leaving us to question their significance in the overall risk model. adORS overcomes this limitation, as evidenced not only by the improved stratification results without APOE, but also by the gene-level interpretability analyses with and without APOE. Within the adORS framework, the genetic contributions are more balanced and dispersed, ensuring that no single gene, including APOE, monopolizes the risk prediction (Supplementary Figures 3A,C). This distribution not only provides a more holistic view of the multifactorial contributors to the disease, but also unmask the latent potential of these genes in AD pathology. The ability of adORS to interpret specific genetic information allows the

focus to be shifted from general relative risk based on a given population to the information that adORS can provide for a single individual, thus permitting the possibility of personalized treatments for AD.

As expected, APOE had the highest contribution in the adORS. with.APOE models for both CN vs. AD classification and MCI conversion prediction (Figures 2A,C; Supplementary Figures 3A,C). APOE pathogenesis is known to affect amyloid- β peptide aggregation and clearance, as well as tau neurofibrillary degeneration, microglia and astrocyte responses, and blood-brain barrier disruption (Serrano-Pozo et al., 2021). ATF6 achieved the subsequent highest contribution in adORS.with.APOE_{CNAD} (Figure 2A; Supplementary Figure 3A). This gene was also the top contributor for adORS_{MCI-CP} (Figure 2D; Supplementary Figure 3D). In a recent study, ATF6 demonstrated an involvement in reducing amyloid precursor protein expressions in AD model mice (Du et al., 2020). ATF6 was also found to be strongly activated in AD and amyotrophic lateral sclerosis (ALS) individuals in pathways related to endoplasmic reticulum homeostasis (Montibeller and de Belleruche, 2018). ING5, a tumor suppressor protein that inhibits cell growth, achieved the second-highest contribution in adORS. with.APOE_{MCI-CP} (Figure 2C; Supplementary Figure 3C). ING5 is associated with various cancer types and rare diseases involving neuronal abnormalities, such as Ohdo Syndrome (Gou et al., 2015; Zhang et al., 2015; Gene Cards Human Gene Database, 2022). In particular, ING5 was found to be highly expressed in stem cells which promotes self-renewal of brain tumor initiating cells, which typically lead to glioblastomas (Wang et al., 2018). Another study has found that ING5 is silenced by short hairpin RNAs during the generation of neuronal precursor cells (Zhang et al., 2013). The evidence of ING5 in neuronal-related development suggests a possible connection to AD pathophysiology.

In adORS_{CNAD}, EFCAB11, which is predicted to enable calcium binding activity, was the highest contributing factor for AD classification (Figure 2B; Supplementary Figure 3B). Circular RNAs (circRNAs), including EFCAB11, have been found to be highly expressed in the brain compared to other tissues, suggesting that circRNAs may constitute as relevant aging biomarkers (Pan et al., 2020). One research group found EFCAB11 to be overexpressed in brain samples of patients with multiple system atrophy, another type of neurodegenerative disorder (Chen et al., 2016). Although there is little information specific to EFCAB11 in AD, studies have found connections of alternative circRNAs to AD, such as dysregulated ciRS-7-miR-7 interaction in the hippocampus of AD patients (Lukiw, 2013; Zhao et al., 2022). Such evidence proposes circRNAs as potential targets for AD treatment.

Other genes observed in the various adORS models and supported by AD-related literature include PTPN5 and VAT1L. PTPN5 (seen in Figure 2; Supplementary Figures 3A–D) is involved in pathways that regulate neuronal signal transduction, neuronal maturation and survival, synaptic function, and learning and memory. Abnormal alterations in PTPN5 activity are associated with several neurological disorders, and it is known that upregulation of PTPN5 may have a causal role in Alzheimer's disease (Chin et al., 2005; Saavedra et al., 2011). VAT1L (seen in Figures 2A,B; Supplementary Figures 3A,B) has been identified as associated with functions related to neuronal maintenance, neurotransmission, and Tau pathology (Jain et al., 1991; Dringen, 2000).

SIK3 and *CD46* displayed reversed directions of effect from adORS_{MCI-CP} to adORS_{CNAD} (Figures 2B,C). In MCI-CP, lower gene burden values for both genes resulted in higher risk scores, whereas this effect was reversed in CNAD. Limited investigation exists on *SIK3*'s role in AD. A gain-of-function *SIK3* mutation has previously been shown to cause an acute decrease in total wake time, resulting in high sleep need despite increased sleep amount (Funato et al., 2016; Wang et al., 2018; Zhou et al., 2022). Interestingly, literature also evidences that sleep disturbances occur more frequently as AD increases in severity. Excessive daytime sleepiness, one of the most common sleep disturbances in demented patients, is characterized by symptoms that align with the *SIK3* mutation, as mentioned above (Vitiello et al., 1990; Honda et al., 2018; Wang and Holtzman, 2020). Further research linking *SIK3*-involved pathways, sleep regulation, and AD may provide additional insight in AD pathology. Similar to *SIK3*, *CD46* is also rarely discussed in relation to AD. *CD46* is a significant complement receptor protein that promotes human cytotoxic CD8+ T cell activity and relates with autoimmune diseases (Liszewski and Atkinson, 2015; Arbore et al., 2018; Stein et al., 2019). One study identified *CD46* as significantly dysregulated in AD patients based on a transcriptome-wide meta-analysis of blood-based microarray gene expression profiles (Nho et al., 2020). The complement system has also been previously linked to AD. An AD-protective variant of *CD33*—another complement protein—is postulated to truncate the protein, reducing functions such as cell signaling and inflammation (Siddiqui et al., 2017). This suggests that lower expression of complement genes may have beneficial effects on the brain, supporting our data that higher *CD46* gene burden values resulted in higher risk scores in the CNAD task. The reversal of effects for both genes may indicate a potential shift in their roles within MCI progression to AD. We suggest further research exploring the effects of *SIK3* and *CD46* particularly in MCI converter patients.

In clinical settings, adORS can be particularly valuable for trials targeting therapies to prevent disease progression. By identifying cognitively normal adults at high AD risk through an elevated adORS, clinicians can concentrate on those most representative of disease progression and most likely to benefit. This approach also enhances trial efficiency and cost-effectiveness (Lewis and Vassos, 2020). adORS also has the potential to aid treatment decisions. For example, individuals with a low adORS might postpone screenings, whereas those with a higher score could start earlier than typically advised (Lewis and Vassos, 2020). Especially for a condition such as AD, in which treatments target symptom management, pinpointing individuals at high genetic risk for earlier interventions can elevate their quality of life. However, it is crucial to note that adORS merely indicates a genetic predisposition to the disease and cannot serve as a diagnosis. As risk scores gain broader applications in both clinical and research settings, recognizing the difference between prediction and definitive diagnosis is essential.

There are several limitations in this study. Note that the success of a gene-level burden analysis depends on several conditions (Li and Leal, 2008). First, the model input should be limited to functional variants and exclude nonfunctional variants. Restricting input to such is crucial to building an interpretable and well-generalized model that performs well across multiple test cohorts. Currently, our model collapses both functional and nonfunctional variants into genes, which likely introduces noise and thus hinders our model from detecting relevant genes with increased accuracy. Although

aggregation methods can amplify association signals and minimize the degrees of freedom more effectively than variant-level tests, they are not always robust to the misclassification of nonfunctional variants. We are actively exploring this avenue in our ongoing research, in which nonfunctional variants are separately considered and grouped according to their intergenic regions. Second, we assume that the direction of effect is consistent within genes. Indeed, two variants with opposite directions of effect will be canceled out. While we adjust for variants negatively associated with AD through the complement operation (see section 2.5), it is also vital to weigh the magnitude and statistical relevance of these effects. As long as we implement a gene-level burden analysis, information loss is inevitable due to the inclusion of nonfunctional variants and exclusion of functional variants. Enhancing the representation of nonfunctional variants is expected to improve the prediction performance. We are investigating the possibility as a sequel to this work.

To ensure generalizability, adORS must be validated on external cohorts. The datasets suitable for training are somewhat restrictive, as both genomic and neuroimaging data must be available to compute our proposed risk score. Despite the constrained number of training data, our model still demonstrates improved performance. Gathering larger training sets is likely to produce a significant boost in accuracy. For validation purposes, adORS has the benefit of requiring only genomic data. For example, by leveraging the adORS generated using ADNI, we can extend its application to other sizable cohorts with available genomic data, such as the Alzheimer's Disease Sequencing Project (ADSP) (2020), and evaluate its proficiency in stratifying the population based on risk levels. Among the many neuroimaging biomarkers that are available in the ADNI dataset, only two were used to generate adORS. In doing so, we may have inadvertently excluded SNPs which are highly associated with other AD biomarkers and may therefore hold predictive ability for AD risk. Additionally, this score only relies on common variants and excludes rare variants. We are considering these limitations as well in the continuations of this study.

Overall, we developed adORS to leverage genetic information that can identify individuals with high or low risk of developing Alzheimer's disease. Compared to conventional PRS, adORS was able to improve prediction performance, risk-based stratification of patients, and interpretability of genetic factors contributing to risk. Further study is required to functionally validate the genes in adORS and their roles in underlying AD-related processes, and ultimately contribute to risk or progression-related biomarkers that can aid in the downstream development of therapeutic treatments for AD.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material. Further inquiries can be directed to the corresponding author.

Ethics statement

The ADNI study was approved by the individual Institutional Review Boards of all participating institutions, with written informed consent obtained from all participants or their guardians according to

the Declaration of Helsinki (consent for research). All data were deidentified.

Author contributions

ES: Conceptualization, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. GL: Conceptualization, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. S-HJ: Formal analysis, Validation, Writing – original draft, Writing – review & editing. ZW: Formal analysis, Writing – original draft, Writing – review & editing. JB: Formal analysis, Writing – original draft, Writing – review & editing. KN: Resources, Writing – review & editing. HH: Resources, Writing – review & editing. CD: Resources, Writing – review & editing. AS: Resources, Writing – review & editing. PT: Resources, Writing – review & editing. LS: Resources, Writing – review & editing. DK: Conceptualization, Funding acquisition, Resources, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the National Institutes of Health (U01 AG068057 and R01 AG071470). Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI; National Institutes of Health, Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd. and its affiliated company

References

- ADNI PET analysis method (2023). Available at: <https://adni.loni.usc.edu/methods/pet-analysis-method/> (Accessed September 28, 2023).
- ADNI-WGS-1 (2020). DSS NIAGADS. Available at: <https://dss.niagads.org/sample-sets/snd10002/> (Accessed October 17, 2020).
- ADSP FUS1 WGS (2021). DSS NIAGADS. Available at: <https://dss.niagads.org/sample-sets/snd10020/> (Accessed October 17, 2022).
- Albert, M. S., DeKosky, S. T., Dickson, D., Dubois, B., Feldman, H. H., Fox, N. C., et al. (2011). The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement.* 7, 270–279. doi: 10.1016/j.jalz.2011.03.008
- Altmann, A., Scelsi, M. A., Shoai, M., de Silva, E., Aksman, L. M., Cash, D. M., et al. (2020). A comprehensive analysis of methods for assessing polygenic burden on Alzheimer's disease pathology and risk beyond APOE. *Brain Commun.* 2:fcz047. doi: 10.1093/braincomms/fcz047
- Alzheimer Association (2021). 2021 Alzheimer's disease facts and figures. *Alzheimers Dement.* 17, 327–406. doi: 10.1002/alz.12328
- Alzheimer's Disease Sequencing Project (ADSP) (2020). DSS NIAGADS. Available at: <https://dss.niagads.org/studies/sa000001/> (Accessed October 18, 2022)
- Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.
- Arbore, G., West, E. E., Rahman, J., Le Fricc, G., Niyonzima, N., Pirooznia, M., et al. (2018). Complement receptor CD46 co-stimulates optimal human CD8(+) T cell effector function via fatty acid metabolism. *Nat. Commun.* 9:4186. doi: 10.1038/s41467-018-06706-z
- Baker, E., and Escott-Price, V. (2020). Polygenic risk scores in Alzheimer's disease: current applications and future directions. *Front. Digit. Health* 2:14. doi: 10.3389/fdgh.2020.00014
- Caminiti, S. P., Ballarini, T., Sala, A., Cerami, C., Presotto, L., Santangelo, R., et al. (2018). FDG-PET and CSF biomarker accuracy in prediction of conversion to different dementias in a large multicentre MCI cohort. *Neuro Image Clin.* 18, 167–177. doi: 10.1016/j.nicl.2018.01.019
- Chaudhury, S., Brookes, K. J., Patel, T., Fallows, A., Guetta-Baranes, T., Turton, J. C., et al. (2019). Alzheimer's disease polygenic risk score as a predictor of conversion from mild-cognitive impairment. *Transl. Psychiatry* 9:154. doi: 10.1038/s41398-019-0485-7
- Chen, B. J., Mills, J. D., Takenaka, K., Bliim, N., Halliday, G. M., and Janitz, M. (2016). Characterization of circular RNAs landscape in multiple system atrophy brain. *J. Neurochem.* 139, 485–496. doi: 10.1111/jnc.13752
- Chin, J., Palop, J. J., Puoliväli, J., Massaro, C., Bien-Ly, N., Gerstein, H., et al. (2005). Fyn kinase induces synaptic and cognitive impairments in a transgenic mouse model

Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnagi.2023.1281748/full#supplementary-material>

- of Alzheimer's disease. *J. Neurosci.* 25, 9694–9703. doi: 10.1523/JNEUROSCI.2980-05.2005
- Cuyvers, E., and Sleegers, K. (2016). Genetic variations underlying Alzheimer's disease: evidence from genome-wide association studies and beyond. *Lancet Neurol.* 15, 857–868. doi: 10.1016/S1474-4422(16)00127-7
- Desikan, R. S., Fan, C. C., Wang, Y., Schork, A. J., Cabral, H. J., Cupples, L. A., et al. (2017). Genetic assessment of age-associated Alzheimer disease risk: development and validation of a polygenic hazard score. *PLoS Med.* 14:e1002258. doi: 10.1371/journal.pmed.1002258
- Dringen, R. (2000). Metabolism and functions of glutathione in brain. *Prog. Neurobiol.* 62, 649–671. doi: 10.1016/S0301-0082(99)00060-X
- Du, Y., Liu, X., Zhu, X., Liu, Y., Wang, X., and Wu, X. (2020). Activating transcription factor 6 reduces Aβ1-42 and restores memory in Alzheimer's disease model mice. *Int. J. Neurosci.* 130, 1015–1023. doi: 10.1080/00207454.2020.1715977
- Foster, N. L., Heidebrink, J. L., Clark, C. M., Jagust, W. J., Arnold, S. E., Barbas, N. R., et al. (2007). FDG-PET improves accuracy in distinguishing frontotemporal dementia and Alzheimer's disease. *Brain* 130, 2616–2635. doi: 10.1093/brain/awm177
- Funato, H., Miyoshi, C., Fujiyama, T., Kanda, T., Sato, M., Wang, Z., et al. (2016). Forward-genetics analysis of sleep in randomly mutagenized mice. *Nature* 539, 378–383. doi: 10.1038/nature20142
- Gene Cards Human Gene Database (2022). ING5 gene—Gene Cards. Available at: <https://www.genecards.org/cgi-bin/carddisp.pl?gene=ING5> (Accessed December 5, 2022).
- Gou, W.-F., Shen, D.-F., Yang, X.-F., Zhao, S., Liu, Y.-P., Sun, H.-Z., et al. (2015). ING5 suppresses proliferation, apoptosis, migration and invasion, and induces autophagy and differentiation of gastric cancer cells: a good marker for carcinogenesis and subsequent progression. *Oncotarget* 6, 19552–19579. doi: 10.18632/oncotarget.3735
- Gouveia, C., Gibbons, E., Dehghani, N., Eapen, J., Guerreiro, R., and Bras, J. (2022). Genome-wide association of polygenic risk extremes for Alzheimer's disease in the UK biobank. *Sci. Rep.* 12:8404. doi: 10.1038/s41598-022-12391-2
- Guerreiro, R. J., Gustafson, D. R., and Hardy, J. (2012). The genetic architecture of Alzheimer's disease: beyond APP. *Neurobiol. Aging* 33, 437–456. doi: 10.1016/j.neurobiolaging.2010.03.025
- Harrison, J. R., Mistry, S., Muskett, N., and Escott-Price, V. (2020). From polygenic scores to precision medicine in Alzheimer's disease: a systematic review. *J. Alzheimers Dis.* 74, 1271–1283. doi: 10.3233/JAD-191233
- Honda, T., Fujiyama, T., Miyoshi, C., Ikkyu, A., Hotta-Hirashima, N., Kanno, S., et al. (2018). A single phosphorylation site of SIK3 regulates daily sleep amounts and sleep need in mice. *Proc. Natl. Acad. Sci. U. S. A.* 115, 10458–10463. doi: 10.1073/pnas.1810823115
- Huang, M., Yang, W., Feng, Q., and Chen, W. (2017). Alzheimer's Disease Neuroimaging Initiative. Longitudinal measurement and hierarchical classification framework for the prediction of Alzheimer's disease. *Sci. Rep.* 7:39880. doi: 10.1038/srep39880
- Huq, A. J., Fulton-Howard, B., Riaz, M., Laws, S., Sebra, R., Ryan, J., et al. (2021). Polygenic score modifies risk for Alzheimer's disease in APOE ε4 homozygotes at phenotypic extremes. *Alzheimers Dement (Amst)*. 13:e12226. doi: 10.1002/dad2.12226
- Jain, A., Mårtensson, J., Stole, E., Auld, P. A., and Meister, A. (1991). Glutathione deficiency leads to mitochondrial damage in brain. *Proc. Natl. Acad. Sci. U. S. A.* 88, 1913–1917. doi: 10.1073/pnas.88.5.1913
- Jung, S.-H., Kim, H.-R., Chun, M. Y., Jang, H., Cho, M., Kim, B., et al. (2022). Transferability of Alzheimer disease polygenic risk score across populations and its association with Alzheimer disease-related phenotypes. *JAMA Netw. Open* 5:e2247162. doi: 10.1001/jamanetworkopen.2022.47162
- Kunkle, B. W., Grenier-Boley, B., Sims, R., Bis, J. C., Damotte, V., Naj, A. C., 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. doi: 10.1038/s41588-019-0358-2
- Lacour, A., Espinosa, A., Louwersheimer, E., Heilmann, S., Hernández, I., Wolfgruber, S., et al. (2017). Genome-wide significant risk factors for Alzheimer's disease: role in progression to dementia due to Alzheimer's disease among subjects with mild cognitive impairment. *Mol. Psychiatry* 22, 153–160. doi: 10.1038/mp.2016.18
- Lee, G., Nho, K., Kang, B., Sohn, K.-A., and Kim, D. (2019). Predicting Alzheimer's disease progression using multimodal deep learning approach. *Sci. Rep.* 9:1952. doi: 10.1038/s41598-019-56802-3
- Leonenko, G., Baker, E., Stevenson-Hoare, J., Sierksma, A., Fiers, M., Williams, J., et al. (2021). Identifying individuals with high risk of Alzheimer's disease using polygenic risk scores. *Nat. Commun.* 12:4506. doi: 10.1038/s41467-021-24082-z
- Leung, Y. Y., Valladares, O., Chou, Y. F., Lin, H. J., Kuzma, A. B., Cantwell, L., et al. (2019). VCPA: genomic variant calling pipeline and data management tool for Alzheimer's disease sequencing project. *Bioinformatics* 35, 1768–1770. doi: 10.1093/bioinformatics/bty894
- Lewis, C. M., and Vassos, E. (2020). Polygenic risk scores: from research tools to clinical instruments. *Genome Med.* 12:44. doi: 10.1186/s13073-020-00742-5
- Li, B., and Leal, S. M. (2008). Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *Am. J. Hum. Genet.* 83, 311–321. doi: 10.1016/j.ajhg.2008.06.024
- Liszewski, M. K., and Atkinson, J. P. (2015). Complement regulator CD46: genetic variants and disease associations. *Hum. Genomics* 9:7. doi: 10.1186/s40246-015-0029-z
- Lu, D., Popuri, K., Ding, G. W., Balachandar, R., and Beg, M. F. (2018). Multimodal and multiscale deep neural networks for the early diagnosis of Alzheimer's disease using structural MR and FDG-PET images. *Sci. Rep.* 8:5697. doi: 10.1038/s41598-018-22871-z
- Lukiw, W. J. (2013). Circular RNA (Circ RNA) in Alzheimer's disease (AD). *Front. Genet.* 4:307. doi: 10.3389/fgene.2013.00307
- Lundberg, S., and Lee, S.-I. A unified approach to interpreting model predictions. arXiv [Preprint]. (2017). doi: 10.48550/arXiv.1705.07874
- Marden, J. R., Mayeda, E. R., Walter, S., Vivot, A., Tchetgen Tchetgen, E. J., Kawachi, I., et al. (2016). Using an Alzheimer disease polygenic risk score to predict memory decline in black and white Americans over 14 years of follow-up. *Alzheimer Dis. Assoc. Disord.* 30, 195–202. doi: 10.1097/WAD.0000000000000137
- Montibeller, L., and de Bellerocche, J. (2018). Amyotrophic lateral sclerosis (ALS) and Alzheimer's disease (AD) are characterised by differential activation of ER stress pathways: focus on UPR target genes. *Cell Stress Chaperones* 23, 897–912. doi: 10.1007/s12192-018-0897-y
- Mullard, A. (2021). Landmark Alzheimer's drug approval confounds research community. *Nature* 594, 309–310. doi: 10.1038/d41586-021-01546-2
- Nho, K., Nudelman, K., Allen, M., Hodges, A., Kim, S., Risacher, S. L., et al. (2020). Genome-wide transcriptome analysis identifies novel dysregulated genes implicated in Alzheimer's pathology. *Alzheimers Dement.* 16, 1213–1223. doi: 10.1002/alz.12092
- Ossenkoppelle, R., Tolboom, N., Foster-Dingley, J. C., Adriaanse, S. F., Boellaard, R., Yaqub, M., et al. (2012). Longitudinal imaging of Alzheimer pathology using [11C] PIB, [18F] FDDNP and [18F] FDG PET. *Eur. J. Nucl. Med. Mol. Imaging* 39, 990–1000. doi: 10.1007/s00259-012-2102-3
- Pan, Y.-H., Wu, W.-P., and Xiong, X.-D. (2020). Circular RNAs: promising biomarkers for age-related diseases. *Aging Dis.* 11, 1585–1593. doi: 10.14336/AD.2020.0309
- Portland Institute for Computational Science (2022). *Alzheimer's Disease Modelling Challenge: Modelling the progression of Alzheimer's disease*. Available at: <http://www.p14cs.org/qt-pad-challenge> (Accessed July 26, 2022).
- Riedel, B. C., Thompson, P. M., and Brinton, R. D. (2016). Age, APOE and sex: triad of risk of Alzheimer's disease. *J. Steroid Biochem. Mol. Biol.* 160, 134–147. doi: 10.1016/j.jsbmb.2016.03.012
- Saavedra, A., Giralt, A., Rué, L., Xifró, X., Xu, J., Ortega, Z., et al. (2011). Striatum-enriched protein tyrosine phosphatase expression and activity in Huntington's disease: a STEP in the resistance to excitotoxicity. *J. Neurosci.* 31, 8150–8162. doi: 10.1523/JNEUROSCI.3446-10.2011
- Serrano-Pozo, A., Das, S., and Hyman, B. T. (2021). APOE and Alzheimer's disease: advances in genetics, pathophysiology, and therapeutic approaches. *Lancet Neurol.* 20, 68–80. doi: 10.1016/S1474-4422(20)30412-9
- Shen, L., and Thompson, P. M. (2020). Brain imaging genomics: integrated analysis and machine learning. *Proc. IEEE Inst. Electr. Electron. Eng.* 108, 125–162. doi: 10.1109/JPROC.2019.2947272
- Siddiqui, S. S., Springer, S. A., Verhagen, A., Sundaramurthy, V., Alisson-Silva, F., Jiang, W., et al. (2017). The Alzheimer's disease-protective CD33 splice variant mediates adaptive loss of function via diversion to an intracellular pool. *J. Biol. Chem.* 292, 15312–15320. doi: 10.1074/jbc.M117.799346
- Sierksma, A., Escott-Price, V., and De Strooper, B. (2020). Translating genetic risk of Alzheimer's disease into mechanistic insight and drug targets. *Science* 370, 61–66. doi: 10.1126/science.abb8575
- Sims, R., Hill, M., and Williams, J. (2020). The multiplex model of the genetics of Alzheimer's disease. *Nat. Neurosci.* 23, 311–322. doi: 10.1038/s41593-020-0599-5
- Sperling, R. A., Aisen, P. S., Beckett, L. A., Bennett, D. A., Craft, S., Fagan, A. M., et al. (2011). Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement.* 7, 280–292. doi: 10.1016/j.jalz.2011.03.003
- Stein, K. R., Gardner, T. J., Hernandez, R. E., Kraus, T. A., Duty, J. A., Ubarretxena-Belandia, I., et al. (2019). CD46 facilitates entry and dissemination of human cytomegalovirus. *Nat. Commun.* 10:2699. doi: 10.1038/s41467-019-10587-1
- Suppiah, S., Didier, M.-A., and Vinjamuri, S. (2019). The who, when, why, and how of PET amyloid imaging in management of Alzheimer's disease—review of literature and interesting images. *Diagnostics* 9:65. doi: 10.3390/diagnostics9020065
- Tan, C. H., Bonham, L. W., Fan, C. C., Mormino, E. C., Sugrue, L. P., Broce, I. J., et al. (2019). Polygenic hazard score, amyloid deposition and Alzheimer's neurodegeneration. *Brain* 142, 460–470. doi: 10.1093/brain/awy327
- Vitiello, M. V., Prinz, P. N., Williams, D. E., Frommlet, M. S., and Ries, R. K. (1990). Sleep disturbances in patients with mild-stage Alzheimer's disease. *J. Gerontol.* 45, M131–M138. doi: 10.1093/geronj/45.4.M131

- Wang, C., and Holtzman, D. M. (2020). Bidirectional relationship between sleep and Alzheimer's disease: role of amyloid, tau, and other factors. *Neuropsychopharmacology* 45, 104–120.
- Wang, Z., Ma, J., Miyoshi, C., Li, Y., Sato, M., Ogawa, Y., et al. (2018). Quantitative phosphoproteomic analysis of the molecular substrates of sleep need. *Nature* 558, 435–439.
- Wang, F., Wang, A. Y., Chesnelong, C., Yang, Y., Nabbi, A., Thalappilly, S., et al. (2018). ING5 activity in self-renewal of glioblastoma stem cells via calcium and follicle stimulating hormone pathways. *Oncogene* 37, 286–301. doi: 10.1038/onc.2017.324
- Ward, A., Tardiff, S., Dye, C., and Arrighi, H. M. (2013). Rate of conversion from prodromal Alzheimer's disease to Alzheimer's dementia: a systematic review of the literature. *Dement. Geriatr. Cogn. Dis. Extra.* 3, 320–332. doi: 10.1159/000354370
- Zetterberg, H., and Bendlin, B. B. (2021). Biomarkers for Alzheimer's disease—preparing for a new era of disease-modifying therapies. *Mol. Psychiatry* 26, 296–308. doi: 10.1038/s41380-020-0721-9
- Zhang, Y., Schulz, V. P., Reed, B. D., Wang, Z., Pan, X., Mariani, J., et al. (2013). Functional genomic screen of human stem cell differentiation reveals pathways involved in neurodevelopment and neurodegeneration. *Proc. Natl. Acad. Sci. U. S. A.* 110, 12361–12366. doi: 10.1073/pnas.1309725110
- Zhang, Q., Sidorenko, J., Couvy-Duchesne, B., Marioni, R. E., Wright, M. J., Goate, A. M., et al. (2020). Risk prediction of late-onset Alzheimer's disease implies an oligogenic architecture. *Nat. Commun.* 11:4799. doi: 10.1038/s41467-020-18534-1
- Zhang, F., Zhang, X., Meng, J., Zhao, Y., Liu, X., Liu, Y., et al. (2015). ING5 inhibits cancer aggressiveness via preventing EMT and is a potential prognostic biomarker for lung cancer. *Oncotarget* 6, 16239–16252. doi: 10.18632/oncotarget.3842
- Zhao, Y., Jaber, V. R., and Lukiw, W. J. (2022). Current advances in our understanding of circular RNA (circRNA) in Alzheimer's disease (AD); the potential utilization of synthetic circRNAs as a therapeutic strategy in the clinical management of AD. *Front Drug Discov.* 2:983030. doi: 10.3389/fddsv.2022.983030
- Zhou, R., Wang, G., Li, Q., Meng, F., Liu, C., Gan, R., et al. (2022). A signalling pathway for transcriptional regulation of sleep amount in mice. *Nature* 612, 519–527. doi: 10.1038/s41586-022-05510-6