Received: 31 July 2024 | Revised: 22 October 2024 | Accepted: 8 November 2024

DOI: 10.1002/alz.14444

Alzheimer's & Dementia®
THE JOURNAL OF THE ALZHEIMER'S ASSOCIATION

RESEARCH ARTICLE

Integrating rare pathogenic variant prioritization with gene-based association analysis to identify novel genes and relevant multimodal traits for Alzheimer's disease

Jixin Cao[1] | Cheng Zhang[2] | Chun-Yi Zac Lo[3] | Qihao Guo[4] | Jing Ding[1] | Xiaohui Luo[1] | Zi-Chao Zhang[1] | Feng Chen[5] | the ZIB Consortium | Tian-Lin Cheng[2,6,7,8] | Jingqi Chen[1,6,9] | Xing-Ming Zhao[1,6,7,9,10] | for the Alzheimer's Disease Neuroimaging Initiative

[1]Department of Neurology, Zhongshan Hospital and Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai, China

[2]Institute for Translational Brain Research, Fudan University, Shanghai, China

[3]Department of Biomedical Engineering, Chung Yuan Christian University, Taoyuan, Taiwan

[4]Department of Gerontology, Shanghai Jiao Tong University Affiliated Sixth People's Hospital, Shanghai, China

[5]Department of Radiology, Hainan General Hospital (Hainan Affiliated Hospital of Hainan Medical University), Haikou, Hainan, China

[6]MOE Frontiers Center for Brain Science, Fudan University, Shanghai, China

[7]State Key Laboratory of Medical Neurobiology, Institutes of Brain Science, Fudan University, Shanghai, China

[8]Institute of Pediatrics, National Children's Medical Center, Children's Hospital, Fudan University, Shanghai, China

[9]MOE Key Laboratory of Computational Neuroscience and Brain-Inspired Intelligence, Fudan University, Shanghai, China

[10]Lingang Laboratory, Shanghai, China

Correspondence

Xing-Ming Zhao, and Jingqi Chen, Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, 220 Handan Road, Shanghai, China.
Email: xmzhao@fudan.edu.cn and jingqichen@fudan.edu.cn

Tian-Lin Cheng, Institute for Translational Brain Research, Fudan University, 131 Dong'an Road, Shanghai, China.
Email: chengtianlin@fudan.edu.cn

Abstract

INTRODUCTION: Increasing evidence has highlighted rare variants in Alzheimer's disease (AD). However, insufficient sample sizes, especially in underrepresented ethnic groups, hinder their investigation. Additionally, their impact on endophenotypes remains largely unexplored.

METHODS: We prioritized rare likely-deleterious variants based on whole-genome sequencing data from a Chinese AD cohort (n = 988). Gene-based optimal sequence kernel association tests were conducted between AD cases and normal controls to identify AD-related genes. Network clustering, endophenotype association, and cellular experiments were conducted to evaluate their functional consequences.

Jixin Cao and Cheng Zhang contributed equally to this study.

Funding information: National Institute on Aging; National Institutes of Health, Grant/Award Number: U19AG024904; National Institute of Biomedical Imaging and Bioengineering; Canadian Institutes of Health Research, Grant/Award Numbers: P30AG10161, P30AG72975, R01AG15819, R01AG17917, R01AG036836, U01AG46152, U01AG61356, U01AG046139, P50 AG016574, R01 AG032990, U01AG046139, R01AG018023, U01AG006576, U01AG006786, R01AG025711, R01AG017216, R01AG003949, R01NS080820, U24NS072026, P30AG19610, U01AG046170, RF1AG057440, U24AG061340; Cure PSP, Mayo and Michael J Fox foundations; Arizona Biomedical Research Commission; Arizona Department of Health Services; National Natural Science Foundation of China, Grant/Award Numbers: T2225015, 62433008, 32200537, 31600826, 32371144; Shanghai Science and Technology Commission Program, Grant/Award Number: 23JS1410100; Key Science and Technology Project of Hainan Province, Grant/Award Number: ZDYF2024SHFZ058; National Key Research and Development Program of China, Grant/Award Numbers: 2023YFF1204800, 2020YFA0712403, 2019YFA0111000; Lingang Laboratory & National Key Laboratory of Human Factors Engineering Joint Grant, Grant/Award Number: LG-TKN-202203-01; Lingang Laboratory, Grant/Award Numbers: LG-GG-202401-ADA010100, LG-GG-202401-ADA050100, LG-QS-202203-08; Shanghai Municipal Science and Technology Major Project, Grant/Award Number: 20JC1419500; Natural Science Foundation of Shanghai, Grant/Award Number: 20ZR1403100; Hainan Academician Innovation Platform Fund; Hainan Province Clinical Medical Center

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

ZIB Consortium: A list of authors and their affiliations appears at the end of the paper.

RESULTS: We identified 11 novel AD candidate genes, which captured AD-related pathways and enhanced AD risk prediction performance. Key genes (*RABEP1*, *VIPR1*, *RPL3L*, and *CABIN1*) were linked to cognitive decline and brain atrophy. Experiments showed *RABEP1* p.R845W inducing endocytosis dysregulation and exacerbating toxic amyloid β accumulation, underscoring its therapeutic potential.

DISCUSSION: Our findings highlighted the contributions of rare variants to AD and provided novel insights into AD therapeutics.

KEYWORDS
Alzheimer's disease, biomarkers, rare variant, the endocytic pathway, whole-genome sequencing

Highlights

- Identified 11 novel AD candidate genes in a Chinese AD cohort.
- Correlated candidate genes with AD-related cognitive and brain imaging traits.
- Indicated *RABEP1* p.R845W as a critical AD contributor in the endocytic pathway.

1 | BACKGROUND

Alzheimer's disease (AD) is a progressive neurodegenerative disorder characterized by cognitive deficits, eventually resulting in dementia and death.[1] As the population ages, the prevalence of AD increases rapidly, making it one of the most serious public health concerns in the world.[1] The genetic etiology of AD is complex, and late-onset AD (LOAD, the common form of AD) has a heritability of 58%–79%.[2] Genome-wide association studies (GWAS) have identified more than 70 loci of common variants related to AD risk, mainly involving amyloid and tau pathways.[3] Nevertheless, there is still a large portion of missing heritability, which may be attributable to large-effect genetic variants, such as rare variants.[4]

Advances in next-generation sequencing and genotyping technologies have made it possible to study large numbers of rare genomic variants in parallel. In an exome microarray study of 37,022 LOAD cases and 48,402 controls performed in 2017, Sims and colleagues found rare variants in Triggering Receptor Expressed On Myeloid Cells 2 (*TREM2*), Phospholipase C Gamma 2 (*PLCG2*), and ABI Family Member 3 (*ABI3*) that were genome-wide significantly associated with AD.[5] More recently, exome sequencing and gene-based testing further discovered significant signals in *ATP8B4* and *ABCA1*.[6] In addition, other sequencing studies have identified multiple other risk genes.[7–10] However, each study seemed to capture only a part of the rare variant landscape of AD, as is expected from the nature of their rare occurrences and the high population heterogeneity.[11] A typical example is *TREM2* p.R47H, a recognized AD risk factor in Europeans but rarely found in East Asian populations.[12,13] These advances emphasized the importance of conducting more independent studies in multiple races, and the demand for novel approaches that bypass the insufficient sample size issue.

China has the highest number of AD patients in the world, but there has been less research on the genetic causes of AD in Chinese cases compared to Caucasians. Previous studies have identified *MLKL*,[14] *C7*,[15] *PDE11A*,[16] and *ACAA1*[17] as new risk genes in the Chinese AD population. However, these studies predominantly focused on early-onset or familial AD. Exploring rare variants in sporadic Chinese AD cases, which have rich genetic diversity, might boost our understanding of the genetic mechanisms of AD in a population-aware manner. Compared with whole-exome sequencing (WES), whole-genome sequencing (WGS) avoids systematic bias from capture methods and offers broader coverage, making it ideal for studying rare variants across the genome. Due to its relatively high costs, which limit the sample size, we opted for a hybrid approach combining prediction-powered variant prioritization and gene-based statistical analysis to maximize efficacy. In addition, we aimed to investigate the extent to which prioritized rare disease-relevant variants contribute to AD-related cognitive and brain imaging traits, as suggested by a recent whole-exome study on AD cases of European origin.[18]

In this report, we performed a WGS analysis on 988 Chinese individuals, including 239 cases with sporadic AD. We designed a variant prioritization pipeline to identify rare likely-deleterious variants across the genome. We then performed cross-ethnic optimal sequence kernel association tests (SKAT-O) to identify novel AD-related genes, whose biological plausibility was rigorously assessed by network analysis, pathway enrichment, and endophenotype association. In addition, in vitro experiments further elucidated the functional consequences of the top candidate variant, which might serve as a promising therapeutic target.

2 | METHODS

2.1 | Study cohort

The study cohort was derived from the Zhangjiang International Brain BioBank (ZIB, https://zib.fudan.edu.cn), a platform specializing

Alzheimer's & Dementia®
THE JOURNAL OF THE ALZHEIMER'S ASSOCIATION

in multimodal research on brain diseases that provides extensive molecular and phenotypic data. Overall, WGS was performed on Deoxyribonucleic acid (DNA) samples from 1018 participants, yielding qualified data from 988 subjects after quality control processes (see later sections). These 988 subjects included 239 individuals with AD, 250 with mild cognitive impairment (MCI), 198 with subjective cognitive decline (SCD), 165 with slight cognitive symptom (SCS), and 136 normal controls (NC).

All participants underwent a battery of cognitive assessments, including the Mini-Mental State Examination (MMSE),[19] the Montreal Cognitive Assessment-Basic (MoCA_B),[20] and the Addenbrooke's Cognitive Examination-III (ACE-III),[21] to assess global cognition. Moreover, standardized neuropsychological tests were administered to assess objective cognitive impairment (CI). These tests covered six measures across three cognitive domains: (1) the Auditory Verbal Learning Test (AVLT) for episodic memory[22]; (2) the Animal Verbal Fluency Test (AFT)[23] and Boston Naming Test (BNT)[24] for language function; and (3) the Shape Trail Test (STT-A and STT-B) for executive function.[25] All tests for each participant were conducted during the same interview session to ensure consistency.

The diagnosis of AD followed the guidelines of the National Institute on Aging-Alzheimer's Association (NIA-AA).[26] For individuals who did not meet the diagnostic criteria for dementia, MCI was diagnosed according to the Jak/Bondi criteria.[27] The criteria required either (1) impaired scores on both measures within the same cognitive domain, or (2) one impaired score in each of three different cognitive domains. Impairment was defined as scoring more than 1 standard deviation (SD) below the age-adjusted normative mean. SCD was defined as essentially normal performance on neuropsychological tests, but with self-reported and concerned memory decline within the last 5 years, based on the criteria proposed by Jessen et al.[28] Patients with SCS were determined according to our previously proposed framework.[29] In this study, individuals with MCI, SCD, or SCS were collectively referred to as CI patients. We recruited cognitively normal elderly adults (age > 60) living in Shanghai between 2019 and 2023 as NC. Additional inclusion criteria for NC were: (1) absence of a disease history or family history of other neurological or psychiatric diseases, such as Parkinson's disease, depression, epilepsy, and neurodevelopmental delay; (2) absence of serious somatic diseases; and (3) having adequate vision and hearing.

The procedures of this study were approved by the Ethics Committee of Shanghai Sixth People's Hospital (approval number: 2019-032). All participants or their legal guardians provided written consent for research projects.

2.2 | Replication cohorts

We employed the sequencing data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) and the Alzheimer's Disease Sequencing Project (ADSP) as two replication cohorts. A detailed description and sample inclusion criteria can be found in Supplementary Methods.

RESEARCH IN CONTEXT

1. **Systematic review**: Our literature review using Google Scholar and PubMed revealed: (1) a substantial missing heritability in Alzheimer's disease (AD) is likely attributed to rare variants; (2) whole-genome studies of AD are less common in East Asia compared to European populations; (3) the effects of rare variants on cognitive functions and brain imaging traits remain largely unknown.
2. **Interpretation**: Our research identified 11 potential AD-related genes in a Chinese cohort, implicating known AD pathways and capturing additional disease risks beyond common variants. Further endophenotype analyses linked these genes with cognitive functions and AD-related brain imaging traits. Moreover, we experimentally validated the role of the top locus, *RABEP1* p.R845W, in the endocytic pathway and Amyloid Beta Precursor Protein (APP) amyloid processing.
3. **Future directions**: Expanding studies to larger, same-ethnic samples could enhance our comprehension of AD in underrepresented populations. Additionally, animal studies could offer deeper insights into the mechanisms of rare variants.

2.3 | WGS variant discovery and quality control

Genomic DNA was extracted from the peripheral blood samples. WGS was performed on the Illumina Novaseq 6000 platform with 150-bp paired-end reads, resulting in a mean sequencing depth of 33.46×. The sequence alignment and variant calling followed the Broad Institute best practices implemented by the Sentieon Genomics software (v202010.02).[30] We then performed Variant Quality Score Recalibration (VQSR) to control base quality and conducted further quality control at both the variant and individual levels. The details of each step are presented in Supplementary Methods and Table S1.

2.4 | Variant prioritization

Single nucleotide variants (SNVs) and small insertions and deletions (INDELs) were annotated using Ensembl Variant Effect Predictor (VEP v109, based on ENSEMBL 109).[31] We considered only variants located on autosomes and defined the following criteria to identify rare likely-deleterious variants: (1) the minor allele frequency (MAF) based on the Genome Aggregation Database (gnomAD, v3.1.2), as well as the subpopulation Minor allele frequency (MAF) for the same race (e.g., East Asian) in gnomAD, were both less than 1%[32]; (2) the MAF based on each research cohort was below 5% to avoid systematic errors; (3) variants annotated by VEP that may affect protein function, including nonsense, splice acceptor/donor, frameshift, and missense variants;

(4) variants labeled as High confidence ("HC") by Loss-Of-Function Transcript Effect Estimator (LOFTEE) (categorized as predicted loss-of-function [LoF]), or missense variants with the Rare Exome Variant Ensemble Learner (REVEL) score > 0.75 or categorized as harmful by at least three out of six tools: Combined Annotation Dependent Depletion (CADD), Sorting Intolerant From Tolerant (SIFT), PolyPhen2, Likelihood Ratio Test (LRT), Mendelian Clinically Applicable Pathogenicity (M-CAP), and MutationTaster (categorized as predicted deleterious missense [DMis]). The predictions of these tools were obtained from dbNSFP4.1a[33,34]; (5) variants located in genes with a Gene Damage Index (GDI) below 13.34 (the recommended cutoff based on the distribution of all disease-causing human genes) or known to be AD-causing, as highly mutated genes in healthy individuals would be less likely to be disease-relevant.[35]

2.5 | Identification of rare variants in the AD core genes

Through a literature review, we compiled a list of 25 AD core genes whose associations with AD risk have been extensively confirmed by whole-genome association studies, sequencing, or in vitro experimental analyses (Supplementary Methods, Table S2). We screened LoF variants in these genes and excluded benign variants that have been reported in ClinVar with at least two stars.[36] Fisher's exact test was used to assess the relative risk of the selected variants in disease groups compared with NC. Ultimately, we excluded the variants carried by NCs across all three cohorts and compiled final lists of variants.

2.6 | Discovery of novel AD-associated genes by SKAT-O

We aggregated the rare likely-deleterious variants for each gene and applied the gene-based SKAT-O to assess their association with AD risk.[37] Separate models were constructed for the LoF and DMis variants, comparing the AD and NC groups within each cohort. The set of covariates consisted of gender, age, age squared, the first five genetic principal components, and the number of apolipoprotein E (*APOE*) $\varepsilon 4$ alleles. Education years were included as covariates only for cohorts where this information was available. The analysis was conducted using the R package SKAT (v2.2.5) with the method set to "SKATO", while other parameters were kept at default settings. Genes with fewer than two variants or with a cumulative minor allele count (MAC) of less than two were excluded from the association analyses. When genes surpassed the nominal significant level ($P < 0.05$), logistic regression models were fitted to estimate effect sizes and directions.

For genes with SKAT-O $P < 0.01$ in the ZIB_AD cohort, additional meta-analyses were performed separately on the ZIB_AD cohort with the two replication cohorts. These analyses were accomplished utilizing the R package MetaSKAT (v0.82) with the "method" parameter set to "optimal".[38] Due to differences in ancestry between the discovery and replication cohorts, we used group-specific MAFs for weight calculation and allowed for heterogenous genetic effects. This was achieved by setting "is.separate" to "TRUE" and "combined.weight" to "FALSE". Ultimately, genes were considered candidates if they showed the same effect direction in both the ZIB_AD cohort and a replication cohort and had a false discovery rate (FDR) of less than 0.05 in the meta-analysis.

2.7 | Polygenic risk scores and random forest models

Details of calculating polygenic risk scores (PRS) and building random forest models are given in Supplementary Methods. Briefly, we employed two external datasets to calculate PRS, including the AD GWAS summary data from the Japanese population to match ethnicity,[39] and a larger GWAS summary data from the Caucasian population.[40]

We developed random forest models to predict the diagnosis of each participant, using the ZIB_AD cohort as the training set. We chose another small AD cohort of 134 participants from ZIB as an independent test set, including 88 AD cases and 46 NC, with no sample overlap with the training set. Random forest models were trained using features such as PRS values, the number of variants in each AD core gene, the number of variants in each AD candidate gene, or a combination of them.

2.8 | Network connectivity and enrichment analysis

We used the human gene connectome (HGC) to define biologically plausible distances between two genes.[41] For each candidate gene, we calculated the mean of their distances to each AD core gene. We then conducted 10,000 permutation tests, each with an equal number of genes randomly chosen from all SKAT-O examined genes, to evaluate whether the average distance between the candidate genes and core genes was lower than expected. Furthermore, we applied hierarchical clustering to all AD core genes and candidate genes, using Ward's minimum variance criterion, based on the pairwise distances between them. Sub-clusters of genes were then obtained based on the distance to the root node.

Gene Ontology (GO) enrichment analysis was performed using the gprofiler2 R package with default parameters, and only GO terms labeled "highlight" were considered.[42]

2.9 | Endophenotype analysis

We conducted the gene-based SKAT-O with cognitive and Magnetic resonance imaging (MRI) measurements as outcomes, employing all available participants in the ZIB_AD cohort, as detailed in Supplementary Methods. For differential expression analysis, we used results from the RNA Sequencing (RNAseq) Harmonization Study, supported by the Accelerating Medicines Partnership-Alzheimer's Disease (AMP-AD) consortium, which uniformly processed RNAseq data from the ROSMAP, Mayo, and MSBB studies (https://github.com/

Sage-Bionetworks/amp-rnaseq). We selected results of the differential expression model that compared the AD group with NC of all genders, defining AD by harmonized criteria such as cognitive scores, Braak staging, and tau pathology.

2.10 | In vitro experiments for the *RABEP1* p.R845W variant

We generated the CBE plasmid used for the *RABEP1* p.R845W mutation, followed by stable cell line generation in HEK293T cells. These cells were cultured, transfected, and selected to isolate clones with the desired mutation for functional assays.

We evaluated cell proliferation using the Cell Counting Kit (CCK)-8 assay, apoptosis and cell cycle via flow cytometry, and protein expression through Western blotting. Immunofluorescence staining was performed to visualize protein localization, and amyloid β (Aβ) levels were quantified using enzyme-linked immunosorbent assay (ELISA). Detailed reagent and protocol specifics are available in the Supplementary Methods.

2.11 | Statistical analysis

Statistical analysis was conducted using R software version 4.2.3. Fisher's exact tests were utilized to compare categorical variables, while for continuous variables, normality was assessed with the Shapiro-Wilk test before comparison. If the data conformed to a normal distribution, *t*-tests were applied; otherwise, the non-parametric Mann-Whitney U tests were conducted. The associations of genes with AD and AD-relevant traits were assessed using SKAT-O, accompanied by logistic regression models to determine effect sizes and directions. The Benjamini-Hochberg procedure was employed to control the FDR for multiple testing corrections. Permutation tests were used to assess the distance of candidate genes from the core genes in the HGC.

2.12 | Data and code availability

The variation data reported in this paper have been deposited in the Genome Variation Map (GVM) in the National Genomics Data Center, Beijing Institute of Genomics, Chinese Academy of Sciences and China National Center for Bioinformation, under accession number GVM000652. The ADSP WES data were acquired under the phs000572.v8.p4 dbGAP study accession number (approved number: 93910-9). The ADNI data can be accessed at https://adni.loni.usc.edu/ upon approval. The Japanese AD GWAS summary data can be downloaded from the National Bioscience Database Center (NBDC) Human Database (Research ID: hum0237.v1). The summary data for European AD GWAS can be downloaded from the GWAS Catalog under the accession number GCST90012878. The expressions of genes in brain tissues are available in the AMP-AD portal (synapse ID: syn14237651). All other data supporting the results of this study can be obtained from the authors upon reasonable request.

The commercial or open-source tools used in this study followed their standard guidelines. In-house scripts and pipelines can be found at https://github.com/ZhaoXM-Lab/AD_SNV.

3 | RESULTS

3.1 | Description of the study cohort

The research cohort, termed the ZIB_AD, was part of the ZIB. Following initial quality control, WGS data from 988 unrelated participants of East Asian descent were further analyzed, including 239 individuals with sporadic AD, 613 with varying degrees of CI (250 with MCI, 198 with SCD, and 165 with SCS), and 136 NC (Methods; Table S1). The diagnosis of AD followed widely accepted and well-validated protocols.[26] NC were cognitively normal, with no major serious illnesses, and were age- and sex-matched to the AD group.

Additionally, we incorporated two replication cohorts to validate initial findings from the ZIB_AD cohort: the ADNI WGS cohort, which had a comparable number of participants to ZIB_AD, and the ADSP WES cohort with a larger scale. Only the majority group (European ancestry) among the two replication cohorts was kept for the analysis (Figure S1). A detailed description of the study and replication cohorts can be found in Table 1.

3.2 | Prioritization of rare likely-deleterious variants

We designed a variant annotation, prediction, and prioritization pipeline to uncover rare and high-impact variants (Figure 1; Methods). Briefly, we focused on protein-coding variants with MAF less than 1%. Multiple in silico tools were then applied to assess their potential deleteriousness. These efforts led to the identification of 100,215 rare likely-deleterious variants within the ZIB_AD cohort, comprising 8555 LoF variants and 91,660 DMis variants. On average, 166.71 DMis variants and 15.57 LoF variants were found per person, with no difference observed between AD/CI cases and NCs (Figure S2). These indicated that rare variants conferring AD risk were limited to specific genes or pathways, rather than exerting a genome-wide collective effect.

To validate the proposed role of rare variants in this disorder, we first examined rare likely-deleterious variants in the 25 well-established AD core genes before further exploration and candidate gene detection (Methods; Table S2). We focused primarily on LoF variants, as they are more damaging and convincing than other types. As expected, rare LoF variants in the known AD core genes exhibited higher prevalence in the AD and CI groups compared to NCs (Figure 2A). Similar trends were observed in the replication cohorts following the same procedure, collectively supporting the involvement of rare and high-impact variants in AD pathology (Figure 2A).

We then excluded variants present in NC across all three cohorts and kept 10 located in five AD core genes, including 4 frameshift, 4 stop-gain, and 2 splicing variants (Figure 2B; Table S3). These variants

TABLE 1 Participants characteristics in this study.

	Discovery cohort			Replication cohorts			
	ZIB_AD WGS (EAS, $n = 988$)			ADNI WGS (EUR, $n = 419$)		ADSP WES (EUR, $n = 10,103$)	
Characteristic	AD (239)	NC (136)	MCI (250) SCD (198) SCS (165)	AD (245)	NC (174)	AD (5,550)	NC (4,553)
Female	61.51%	61.03%	66.80% 70.20% 73.33%	40.00%	51.72%	57.40%	58.67%
Age	71.77 (7.45)	70.02 (5.41)	67.91 (7.14) 65.74 (7.08) 65.58 (7.85)	74.15 (7.16)	74.36 (5.61)	75.36 (8.50)	86.56 (3.57)
Edu years	9.87 (4.27)	12.27 (3.68)	10.94 (3.43) 11.74 (2.87) 12.39 (3.09)	15.96 (2.84)	16.55 (2.65)		
MMSE	16.88 (5.08)	28.06 (1.89)	26.31 (2.08) 27.03 (2.00) 27.70 (1.52)	19.05 (6.30)	28.81 (1.49)		
APOE4	45.61%	19.85%	22.40% 19.19% 21.82%	61.22%	22.99%	42.56%	14.15%

Note: Data are sourced from the ZIB_AD, the ADNI, and the ADSP. Each cell represents percentage or mean (standard deviation).
Abbreviations: AD, Alzheimer's disease; ADSP, Alzheimer's disease sequencing project; ANDI, Alzheimer's disease neuroimaging initiative; APOE4, apolipoprotein E ε4 allele carrier; EAS, East Asian; EUR, European; MCI, mild cognitive impairment; MMSE, Mini-Mental State Examination; SCD, subjective cognitive decline; SCS, slight cognitive symptom; WES, whole-exome sequencing; WGS, whole-genome sequencing; ZIB AD, Zhangjiang international brain biobank AD cohort.

were found in 10 AD cases (4.18%) within the ZIB_AD cohort, with more than half (six) being *APOE ε4* negative, suggesting these variants contribute independently of the known AD risk *APOE ε4* allele (Figure 2B). Two of these variants (*ABCA7* p.E919* and *PSEN1* c.869-2A > G) were previously reported to be pathogenic for AD (Table S3). For instance, the LoF variant *PSEN1* c.869-2A > G was found in European AD cases and carried by a 50-year-old female with AD in our cohort.[43,44] This variant, located in the splice acceptor site in intron 8, has been shown to cause the skipping of exon 9, resulting in an aberrant exon 8-10 junction (Figure 2C).[43] Interestingly, despite the proband's relatively young age, brain MRI revealed a more severe right hippocampal atrophy than other AD cases, consistent with a previous study linking *PSEN1* mutations to hippocampal atrophy (Figure 2D).[45] Notably, 5 of the 10 prioritized LoF variants were absent in gnomAD non–East Asian populations, indicating they are ultra-rare and possibly population-specific (Table S3). In summary, our variant prioritization pipeline demonstrated the potential contribution of rare variants to AD.

3.3 | Cross-cohort gene-based SKAT-O revealed 11 AD candidate genes

With AD core genes representing only a small portion of the genome and serving as proof of concept, our next goal is to identify AD candidate genes more likely to be affected by rare likely-deleterious variants in AD than in controls. As described earlier, it is challenging to attain sufficient statistical power in a medium-sized population. To address this, we adopted a two-step approach: First, we performed gene-based SKAT-O between AD and NC groups in the ZIB_AD cohort to identify genes with suggestive significance. Next, we integrated additional cohorts, such as ADNI and ADSP, to conduct a meta-analysis for these genes to identify significant candidates. Independent models were developed for LoF and DMis variants, as detailed in the **Methods**.

In total, 90 genes reached a suggestive significance threshold in the ZIB_AD cohort (SKAT-O $P < 0.01$) (Figures S3–S4; Table S4). Among these, 11 genes passed the significant threshold in a meta-analysis with at least one replication cohort (MetaSKAT-O FDR < 0.05 and in the same direction as in ZIB_AD), thus being referred to as candidate genes (Figure 2E; Table S5). When performing a larger meta-analysis by combining data from all three cohorts, 8 of the 11 candidate genes, with the exception of *TLX3*, *AARSD1*, and *SMAD6*, remained significant (FDR < 0.05), demonstrating the robustness of our findings (Figure 2E; Table S5).

Among the candidate genes, *RABEP1* and *CRADD* have been previously reported as AD risk genes in prior GWAS studies, while *SMAD6* and *AHCTF1* have been associated with CI, as identified through GWAS Catalog.[46] The remaining seven genes are novel findings in this study. *HNMT*, one of these novel genes, emerged as the top-ranked gene in the meta-analysis (MetaSKAT-$O_{ZIB\_AD\text{-}ADNI}$ FDR $= 1.08 \times 10^{-5}$; Figure 2E). This gene encodes a histamine-metabolizing enzyme essential for the regulation of the histaminergic system in the central nervous system (CNS).[47] Interestingly, this enzyme is also an inhibitory target of an early AD drug, tacrine.[48] Previous studies have demonstrated that

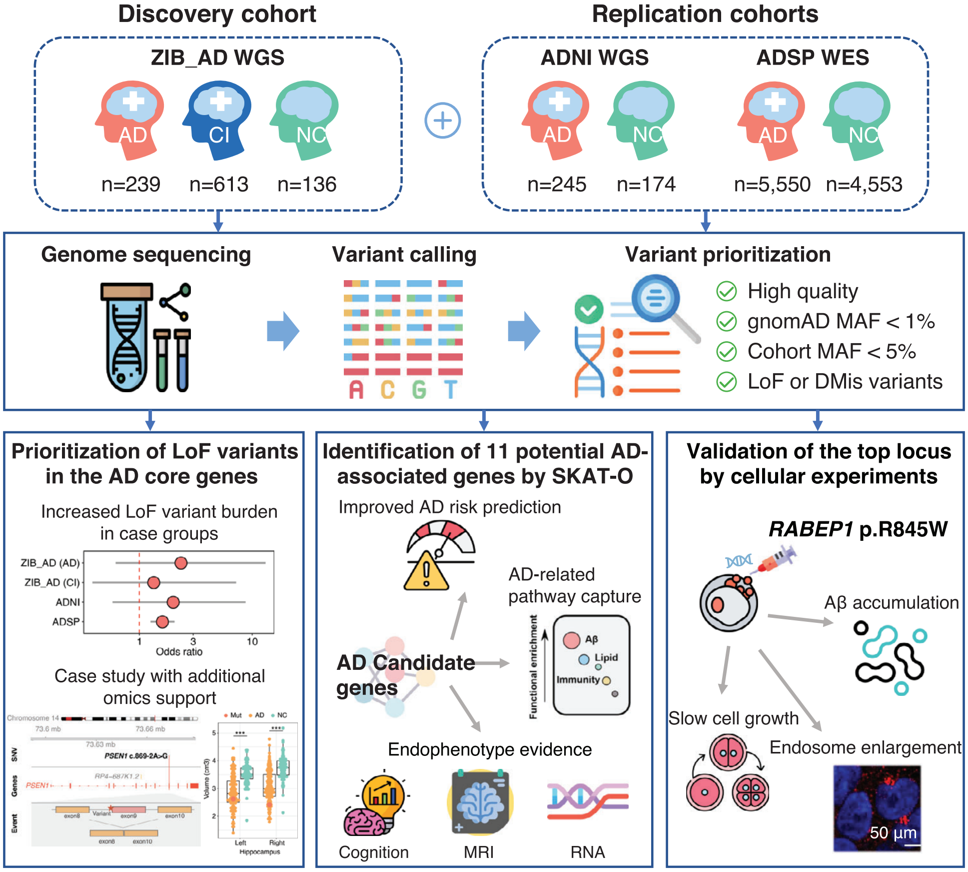

FIGURE 1 The overall study design. WGS analysis was conducted on a Chinese cohort of 988 individuals. European participants from the ADNI and the ADSP cohorts were included as replication samples. Comprehensive variant prioritization strategies were employed to identify rare likely-deleterious protein-coding variants. First, as a proof of concept, an initial screening evaluated the rare LoF variants in 25 known AD core genes. Next, cross-cohort SKAT-O between the AD and NC groups identified 11 candidate genes. The associations with AD were further supported by AD classification models, biological network analyses, and endophenotype analyses. Finally, cellular experiments validated the functional impact of a top locus in AD. AD, Alzheimer's disease; ADNI, Alzheimer's Disease Neuroimaging Initiative; ADSP, Alzheimer's Disease Sequencing Project; CI, cognitive impairment; DMis, predicted deleterious missense; LoF, predicted loss-of-function; MAF, minor allele frequency; NC, normal control; SKAT-O, optimal sequence kernel association test; WGS, whole-genome sequencing.

variants in *HNMT* are associated with neurological disorders, such as intellectual disability and Parkinson's disease.[49,50] Although there were limited reports regarding their association with AD, post mortem studies revealed elevated *HNMT* mRNA expression in the prefrontal cortex of female AD cases.[51] Here, we identified two DMis variants in *HNMT* within the ZIB_AD cohort, carried only by NCs. The DMis variants in *HNMT* were also significantly depleted in AD in both replication cohorts. These findings therefore supported the involvement of neuronal histamine in AD pathology and provided possible protective loci in *HNMT*.[52]

Another top gene, *RABEP1*, stood out among genes with nonprotective loci (MetaSKAT-O$_{\text{ZIB_AD-ADSP}}$ FDR $= 1.18 \times 10^{-3}$; Figure 2E). Rare DMis variants of this gene showed an increased burden in the AD group in our cohort as well as the two replication cohorts. As a crucial gene in the early endosome, *RABEP1* encodes a Rab5 effector protein and plays a role in endocytic membrane docking and fusion.[53] It is widely expressed in numerous tissues, particularly in the nervous system.[54] Interestingly, *RABEP1* was nominated as one of the causal genes in a previous study integrating AD GWAS and myeloid genomes, but the functional mechanism in AD remained unclear.[55] In this sense, the identification of rare variants provided new perspectives on the involvement of this gene in AD.

Overall, we employed cross-cohort SKAT-O to identify 11 potentially AD-associated genes, with *HNMT* and *RABEP1* as the top protective gene and causative gene, respectively.

3.4 | Enhanced capture of AD risks through rare variants in the candidate genes

To demonstrate the robustness of the association between the candidate genes and AD, we evaluated their potential to distinguish AD cases from NCs in an independent Chinese AD cohort using random forest models (Methods). Prior to this, we quantified the effects of common variants using PRS as basic features of the models. Consistent with previous studies, the PRS-driven model showed modest clas-

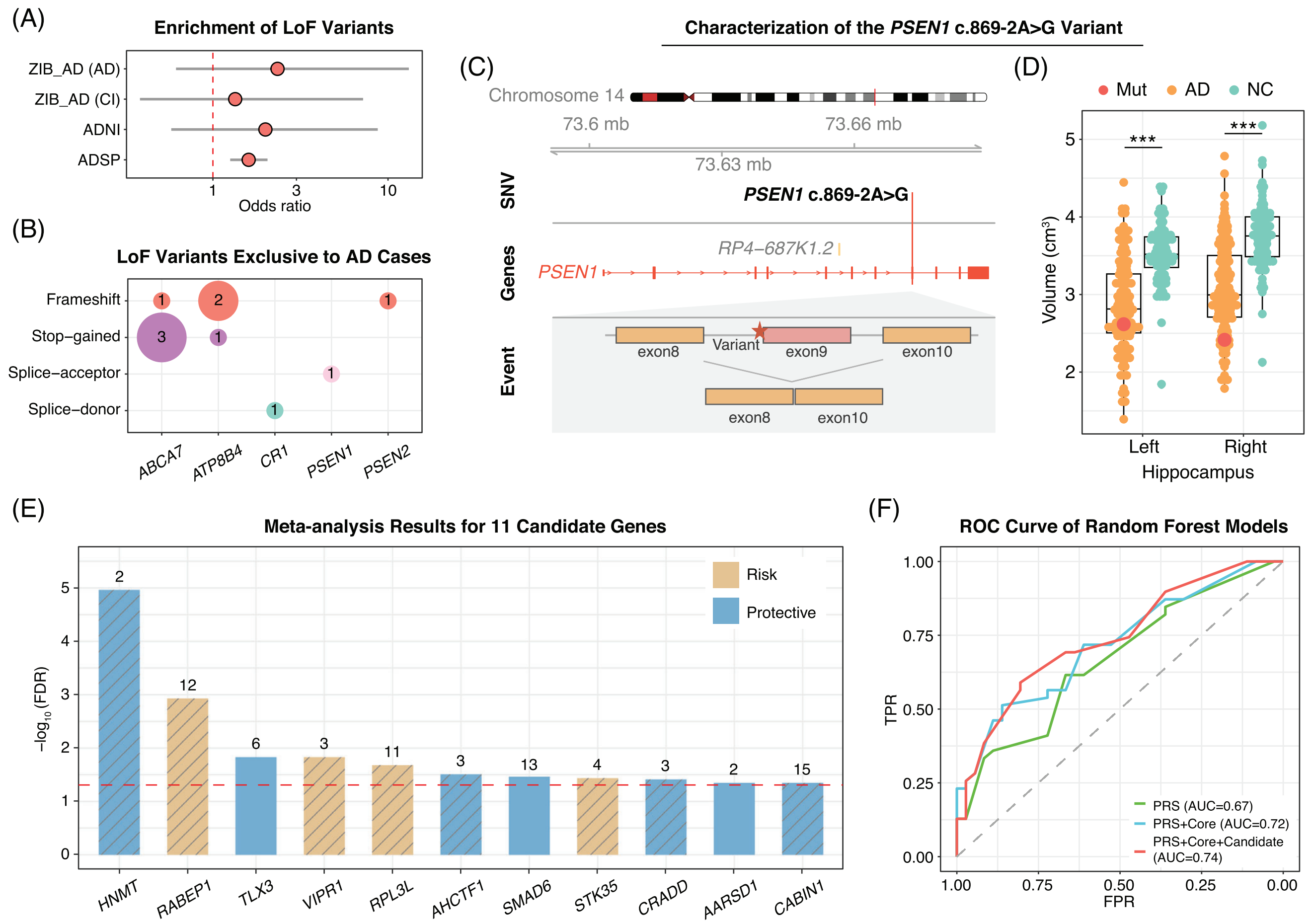

FIGURE 2 Rare likely-deleterious variants in the AD core and novel genes. (A) The burden of rare LoF variants in the AD core genes in case groups compared to NC. Midpoints and error bars represent the OR and their 95% confidence intervals. The dashed vertical red line indicates an OR of 1. (B) The distribution of different types of LoF variants exclusive to AD cases across AD core genes. Each bubble's position reflects the presence of a specific mutation type in a gene. The size of each bubble corresponds to the number of mutations observed, with the numerical values inside providing the exact count of mutations for each gene-mutation type combination. (C,D) Example of the *PSEN1* gene splice-site LoF variant c.869-2A > G. (C) This variant alters the basic region at the 3′ boundary of intron 8, resulting in the skipping of exon 9 and introducing an aberrant exon 8–10 junction. (D) The carrier of this variant was only 50 years old but suffered from unusually severe hippocampal atrophy. The box spans the first to third quartiles; the whiskers extend 1.5 times the interquartile range; and the middle line represents the median. *P* values between the AD and NC groups were calculated using Mann-Whitney U tests. The red dots represent this specific carrier. (E) The MetaSKAT-O results for the 11 novel candidate genes. Each bar represents the negative logarithm MetaSKAT-O FDR for each gene, with the MACs annotated on top. The color of the bar indicates the direction of each gene's effect. The eight genes that remained significant in the three-cohort meta-analysis are highlighted by gray diagonal stripes. The red dashed horizontal line indicates an FDR of 0.05. (F) Performance of random forest models used to distinguish AD cases from NC. Models with different types of variant sets as features are represented by differently colored receiver operating characteristic curves. $^{***}p < 0.001$. AD, Alzheimer's disease; AUC, area under the receiver operating characteristic curve; FDR, false discovery rate; LoF, loss-of-function; MACs, minor allele counts; Mut, mutation carrier; NC, normal controls; OR, odds ratio; SKAT-O, sequence kernel association tests.

sification ability for AD (Figure 2F).[56,57] Interestingly, performance improved when incorporating rare likely-deleterious variants as features, indicating that such rare variants carried risks not accounted for by common loci (Figure 2F). In particular, the inclusion of rare likely-deleterious variants in the candidate genes, in addition to those in the core genes, as features yielded the best performance, underscoring their potential to offer additional information beyond known AD risk factors (Figure 2F). In summary, our findings demonstrated that the candidate genes identified through rare variants could indeed capture risks for AD that were not explained by previously reported risk signals.

3.5 | Highlighted AD-relevant pathways and endophenotypic traits

To further investigate the functional plausibility of the candidate genes in AD mechanisms, we assessed their biological connections with the AD core genes using the HGC.[41] As expected, the candidate genes showed significantly shorter average distances from the core genes compared to random sampling (10,000 times permutation test, $P = 2.19 \times 10^{-2}$; Figure 3A). Furthermore, hierarchical clustering based on pairwise distances revealed that these genes were well-mixed with

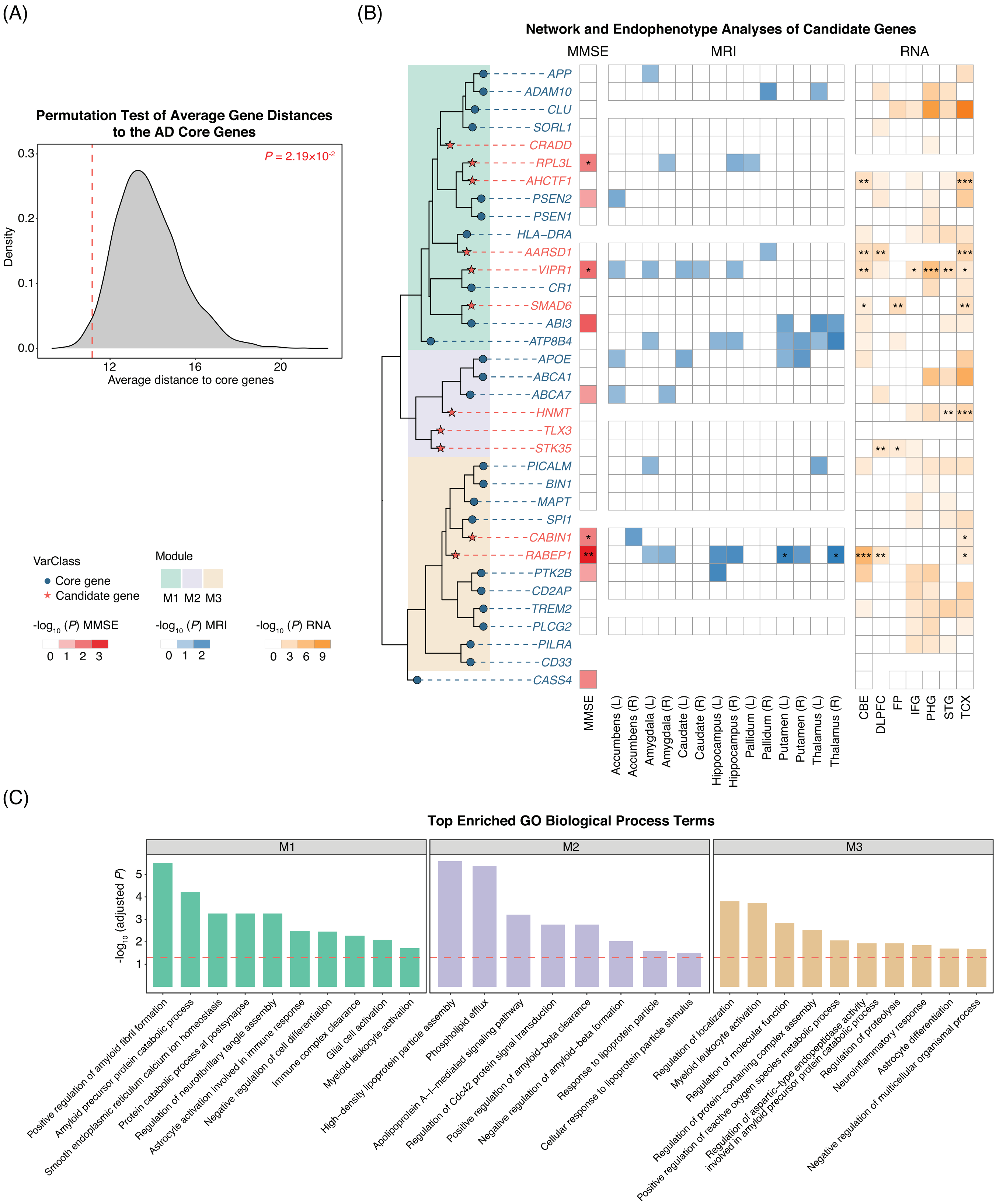

FIGURE 3 Validation of the AD candidate genes via network and endophenotype analysis. (A) The density plot shows the average distances from AD core genes for all genes, tested across 10,000 permutations; the vertical dashed line denotes the average distances between the 11 AD candidate genes and the AD core genes. (B) Hierarchical clustering (left) of AD candidate genes (red) and core genes (blue) according to HGC. Different background colors represent gene modules divided based on hierarchical clustering. Heatmaps illustrate the negative logarithm of *p* values for the gene-based association tests of MMSE scores, subcortical volumes, and differential gene expression across multiple brain regions.

the core genes, resulting in three gene modules (Figure 3B). Notably, all three modules were significantly enriched in established AD pathways such as Aβ metabolism (M1, M2, and M3), lipid metabolism (M2), and neuroinflammation (M1 and M3) (Figure 3C).[58] These observations underscored a recapitulation of known AD mechanisms through novel rare likely-pathogenic variants, reinforcing the biologically significant relationships between the candidate genes and AD.

We then examined the associations of these genes with AD-relevant endophenotypic traits, including cognitive performance and brain volumetric MRI measurements, using multi-modal data available for the ZIB_AD cohort (Methods). As shown in Figure 3B, SKAT-O revealed that four candidate genes (*RPL3L*, *VIPR1*, *CABIN1*, *RABEP1*) were significantly associated with changes in the MMSE score (FDR < 0.05), surpassing random expectations (Fisher's exact test, $P = 2.46 \times 10^{-3}$). Among these genes, *RABEP1* and *VIPR1* exhibited the most significant detrimental effects on cognition (Figure 3B).

We employed the same approach to assess the association of the candidate genes with the volume of subcortical regions. Notably, *RABEP1*, *VIPR1*, and *RPL3L* showed nominally significantly associated with the reduced volumes of multiple regions, including the hippocampus and amygdala, which are well-recognized imaging markers of AD (Figure 3B).[59,60] In addition, carriers of rare likely-deleterious variants in *RABEP1* were accompanied by severe putamen and thalamus atrophy, which was also observed in AD patients in a previous study (Figure 3B).[61]

We evaluated the expression of the candidate genes in various brain tissues using transcriptomic data from the AMP-AD consortium (Methods).[62–65] A total of seven candidate genes were dysregulated (FDR < 0.05) in AD cases in at least one brain region compared to NC, especially in the parahippocampal gyrus (PHG), temporal cortex (TCX), and superior temporal gyrus (STG) (Figure 3B). These brain regions are involved in learning and memory and correspond to the early stages of AD pathology.[66] Taken together, these convergent findings based on the aforementioned multimodal information further supported the potential roles of the candidate genes in AD pathogenesis.

3.6 | Endosomal dysfunction and accumulation of neurotoxic Aβ induced by *RABEP1* p.R845W

Given the positive contribution of the candidate genes to AD pathophysiology, we decided to explore individual loci with large effect sizes, which might serve as promising intervention targets. Across the 11 candidate genes, we identified 52 rare likely-deleterious variants in the ZIB_AD cohort, 8 (15.38%) of which were also found in the replication cohorts (Table S6). Among the other 44 variants, 24 were found to be exclusive to East Asians, according to gnomAD.

Notably, *RABEP1* p.R845W was present in the largest number of AD cases (five) but not in any NC (Table S6). Interestingly, *RABEP1* was also highlighted in our previous association and endophenotype analyses. In addition, the p.R845W variant was absent in the ADSP and ADNI cohorts of non-East Asian participants, although both cohorts did have other *RABEP1* variants.

The p.R845W variant alters from hydrophilic arginine to hydrophobic tryptophan, accompanied by a loss of charge, which is a large change and highly unfavored in terms of conserved amino acid properties. To validate the potential pathological contributions of *RABEP1* p.R845W to AD-related phenotypes, we established cell models carrying *RABEP1*$^{R845W/+}$ in human embryonic kidney (HEK293T) cell lines by base editing (Methods). Surprisingly, the proliferation ability of *RABEP1*$^{R845W/+}$ cells decreased significantly compared to wild-type cells (Figure 4A). We further observed significantly increased early and late apoptosis in *RABEP1*$^{R845W/+}$ cells with the Annexin V-FITC apoptosis assay by flow cytometry (Figure 4B,C). Additionally, cell cycle analysis revealed G2 arrest in *RABEP1*$^{R845W/+}$ cells (Figure S5). Taken together, our results confirmed that *RABEP1* p.R845W was detrimental to cell proliferation, which might be attributed to enhanced cell apoptosis and dysregulated cell cycle processes.

Endosomal defects have been proven to be one of the most typical cellular phenotypes of AD.[67] As *RABEP1* is involved in endocytic membrane docking and fusion, we asked whether the p.R845W variant interfered with endosomal processes. Western Blot showed that expression of Rab5, a key protein in the early endosome, was significantly upregulated in *RABEP1*$^{R845W/+}$ cells (Figure 4D,E; Figure S6). Other key proteins, Rabex-5 and Rabaptin-5, the latter encoded by *RABEP1*, also showed upregulation trends (Figure 4D,E; Figure S6). Notably, immunofluorescence of early endosome antigen 1 (EEA1) revealed that the diameter of the EEA1-positive early endosome was increased in *RABEP1*$^{R845W/+}$ cells (Figure 4F,G). The enlarged endosome meant endosomal dysfunction,[68] which is consistent with the early manifestations of AD.[67,68]

The early endosome serves as the first major sorting station for Amyloid Beta Precursor Protein (APP) and the primary site of Aβ peptide generation.[69] We hypothesized that mutation-induced endosomal abnormalities would affect Aβ metabolism. As expected, enzyme-linked immunosorbent assays showed elevated Aβ42 and a moderately reduced Aβ40 in *RABEP1*$^{R845W/+}$ cells as compared to wild-type cells, resulting in significantly increased ratios of Aβ42 to Aβ40 (Figure 4H,I). These findings indicated that the *RABEP1* p.R845W mutation inter-

Only tests with unadjusted *p* values less than 0.05 are colored and indicated by varying degrees of transparency. Asterisks indicate the significance levels of associations after controlling for FDR in candidate genes. Borderless blocks signify either missing data or tests that do not meet the SKAT-O test criteria. (C) The top significantly enriched biological process terms for each gene module, displaying up to the top 10. ***FDR < 0.001; ***FDR < 0.01; *, FDR < 0.05. AD, Alzheimer's disease; CBE, cerebellum; DLPFC, dorsolateral prefrontal cortex; FDR, false discovery rate; FP, frontal pole; HGC, human gene connectome; IFG, inferior frontal gyrus; MMSE, Mini-Mental State Examination; PHG, parahippocampal gyrus; STG, superior temporal gyrus; SKAT-O, sequence kernel association tests; TCX, temporal cortex.

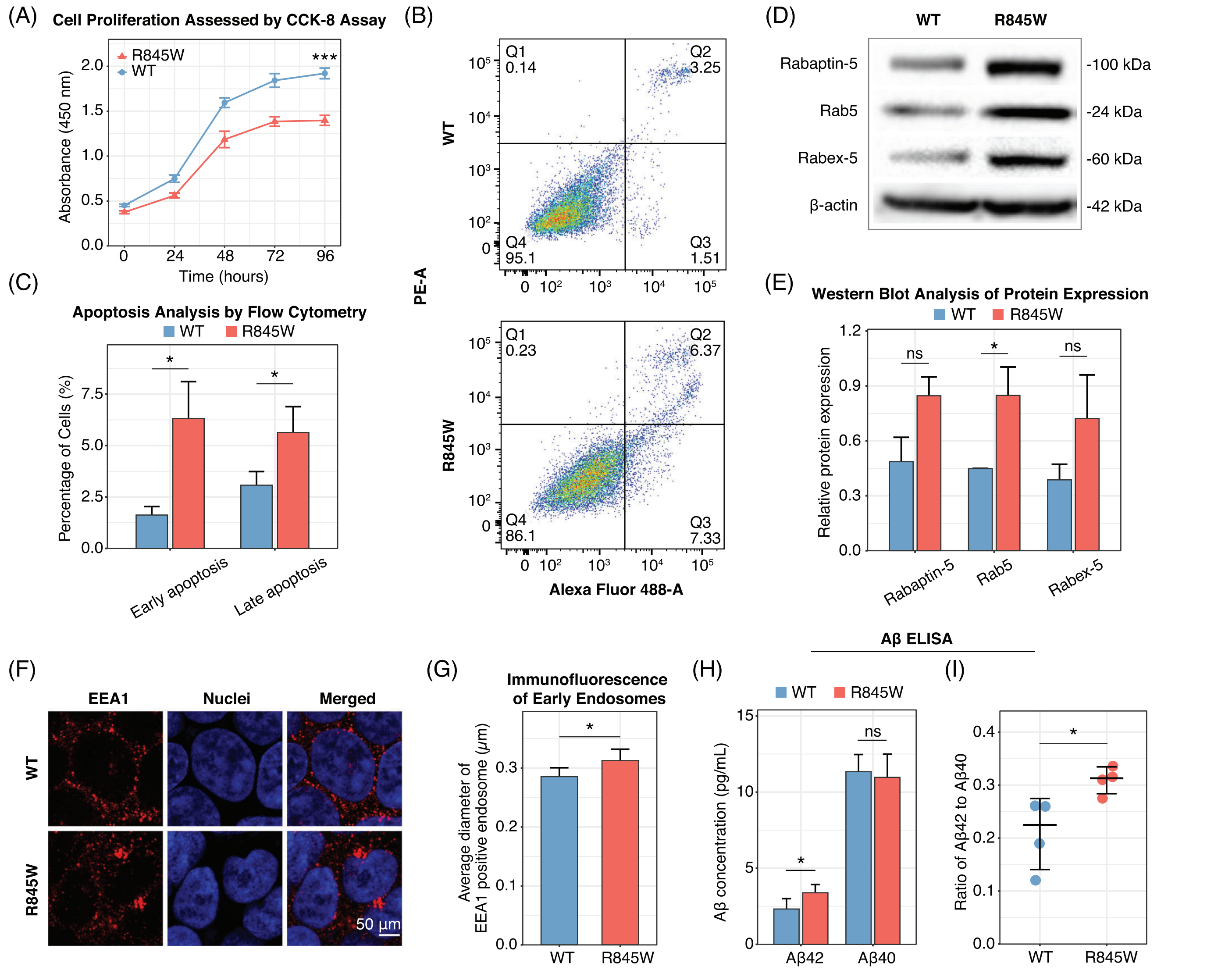

FIGURE 4 Functional validation of the *RABEP1* p.R845W variant. (A) The proliferative capacity of cells measured by the CCK8 assay. (B,C) Images (B) and quantification (C) of the apoptosis assay by flow cytometry. (D,E) Western blot analysis (D) and quantification (E) in *RABEP1*$^{R845W/+}$ cells and wild-type HEK293T cells. (F,G) Immunofluorescence (F) and quantification (G) of EEA1-positive early endosomes. Representative immunofluorescence images were stained with EEA1 (red) and nuclei (blue). The diameters of EEA1-positive endosomes were quantified by fluorescence intensity. Scale bar, 50 µm. (H,I) Concentration of Aβ42 and Aβ40 (H), and the ratio of Aβ42 to Aβ40 (I) detected by ELISA. *P* values were calculated using unpaired two-sample *t*-tests if the data conformed to a normal distribution; otherwise, the non-parametric Mann–Whitney U test was used. $^{*}p < 0.05$; $^{**}p < 0.01$; ns, not significant; CCK8, Cell Counting Kit-8; WT, wild-type HEK293T cells; R845W, HEK293T cells carrying *RABEP1*$^{R845W/+}$.

fered with the APP pathway and led to the abnormal accumulation of neurotoxic Aβ.[70]

Overall, we experimentally validated the AD-related cellular phenotypes of *RABEP1* p.R845W and highlighted that endocytosis dysregulation might be an important causative factor for AD.

4 | DISCUSSION

In the present study, we depicted the rare variant landscape of the sporadic Chinese AD cases, revealing increased burdens of rare LoF variants in the AD core genes. We nominated 11 candidate genes and reinforced their association with AD through network clustering, endophenotype analyses, and gene expression. Our prediction model showed that rare variants could capture additional AD risks not accounted for by previously established common loci. Additionally, we provided *RABEP1* p.R845W as a potential therapeutic target, underscoring the significance of the endocytic pathway in AD. Our findings gave new insights into the genetic mechanisms of AD and emphasized that population-specific rare variants could enrich the understanding of the AD molecular basis.

The screening for rare likely-deleterious variants enriched our comprehension of AD inheritance. We observed significant enrichment of LoF variants in the AD core genes across all cohorts, despite these genes being predominantly identified in studies of Caucasian populations. Notably, rare variants demonstrated promising predictive capabilities for AD risk beyond PRS, implying their ability to capture disease risks beyond common variants. Given the modest sample size of our study, we opted to collapse rare variants into the gene level to increase statistical power and conducted meta-analyses for validation. These approaches enabled us to identify novel candidate genes enriched in well-established AD pathways, such as amyloid processing and tau pathology, reinforcing the biological relevance of our findings. In the meta-analysis of all three cohorts, we observed reduced significance for a few candidate genes, which exhibited different effect directions across cohorts. This was driven primarily by the low rate of shared variants rather than differences in the frequencies of the same variants across populations. This emphasizes that the low frequency of rare variants may lead to sampling differences, as these variants are often underrepresented or entirely absent in certain cohorts. Despite these cohort-specific variations, the convergence of rare variants into shared genes and pathways suggests that rare variants might contribute to AD risk through converging biological mechanisms across populations.

The identification of candidate genes offered promising new insights for AD genetics. Notably, *RABEP1* stood out in association analyses and in vitro experiments, underscoring the significant role of endocytosis in AD. Endocytic organelles are the primary sites of APP amyloid processing,[53,67] where APP is sequentially cleaved by β- and γ-secretases, ultimately forming toxic Aβ.[71] Previous studies observed an increased burden of rare variants in the endocytic pathway in AD cases.[72,73] Our findings reinforced this conclusion and nominated *RABEP1* p.R845W as a possible causal locus. *RABEP1* acts as an effector of the small GTPase Rab5 in early endosomes, the major amyloid processing site for APP.[53,67] The increased Aβ42 to Aβ40 ratio in *RABEP1*$^{\mathrm{R845W/+}}$ cells indicated the variant-induced excessive amyloid cleavage of APP and the accumulation of toxic Aβ peptides. Interestingly, the p.R845W variant was highly detrimental to cell proliferation and survival and induced expanded early endosomes. These strikingly abnormal cellular phenotypes, along with the upregulation of Rab5 protein, might indicate the continuous overactivation of Rab5, as demonstrated in previous studies.[74,75] Notably, multiple AD core genes (i.e., *BIN1*, *SORL1*, and *PICALM*) were also involved in the endocytic pathway and harbored rare likely-deleterious variants in the ZIB_AD cohort, further supporting the potential of the endocytic pathway as a therapeutic target.[76–78]

Functional analysis of candidate genes encouraged us to investigate previously less-studied AD mechanisms. Rare variants in *HNMT*, an essential gene for histamine degradation, showed consistent protective effects against AD in all three cohorts. Histamine functions as a neurotransmitter in the CNS that regulates learning, memory, cognition, and motor functions, all of which are severely impaired in AD.[79,80] Moreover, AD patients exhibited alterations in the histaminergic system, such as reduced histamine levels in various brain regions.[81] Therefore, extensive efforts have been undertaken to increase brain histamine levels by inhibiting *HNMT* as a novel approach for AD treatment.[82] In this sense, the identification of AD-protective loci in *HNMT* could guide the development of relevant inhibitors.

Consistent with previous studies,[83] although we observed relatively similar patterns at the gene and pathway levels across three cohorts, high-impact rare variants were less frequently shared between Chinese and European AD populations, no matter in the AD core genes or candidate genes. This suggests that the specific rare variants driving gene-disease associations may differ across populations, likely due to the genetic heterogeneity of AD and differences in population genetic structure.[4] Nevertheless, our results highlight the value of studying AD in the underrepresentation populations, which may provide novel candidate variants, genes, and mechanistic insights for the disease.

The current research also has some limitations. Given the genetic heterogeneity of AD and the low statistical power of rare variants, the current sample size is insufficient to detect much rarer disease-associated genes. Additionally, the underrepresentation of East Asian populations in AD studies also limited the cross-cohort replication of variants.[83] This emphasizes the need for larger, more ethnically homogenous cohorts to validate these associations more reliably. It is also worth noting that rare variants located in non-coding regions may also bear disease risk. Nevertheless, deciphering their functionality and risk contribution is still challenging, necessitating more robust methods and large-scale screening experiments. Moreover, other forms of variants, such as structural variants, could also contribute to AD and warrant further investigation.

In conclusion, we explored the role of rare likely-deleterious variants in a Chinese AD population, nominated promising candidate genes, and validated a top candidate through base editing and functional experiments. These findings may contribute to further understanding of the genetic mechanisms of AD and provide new insights into the development of targeted therapies.

AUTHOR CONTRIBUTIONS

Xing-Ming Zhao conceived and supervised the project. Jingqi Chen and Tian-Lin Cheng designed and supervised the computations and the experiments, respectively. Jixin Cao conducted data processing and analysis. Cheng Zhang conducted the in vitro experiments. Xiaohui Luo and Zi-Chao Zhang performed parts of the endophenotypic analyses. Jixin Cao wrote the manuscript. Xing-Ming Zhao, Jingqi Chen, Tian-Lin Cheng, and Jing Ding helped revise the manuscript. Qihao Guo, Chun-Yi Zac Lo monitored the collection of samples and the construction of the cohort. Feng Chen contributed to the revision process. All authors read and approved the final version of the manuscript.

ACKNOWLEDGMENTS

We express our appreciation to all the original data sources and the individuals who consented to providing samples. We also acknowledge Dr. Liting Song and Dr. Yuchen Liu for their guidance in the pre-processing of transcriptomes and brain images. Data collection and sharing for the ADNI is funded by the National Institute on Aging (National Institutes of Health Grant U19AG024904). The grantee organization is the Northern California Institute for Research and Education. In the past, ADNI has also received funding from the National

Institute of Biomedical Imaging and Bioengineering, the Canadian Institutes of Health Research, and private sector contributions through the Foundation for the National Institutes of Health (FNIH) including generous contributions from the following: AbbVie, 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 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 differential expression analysis is based on data obtained from the AD Knowledge Portal (https://adknowledgeportal.org). Data generation was supported by the following NIH grants: P30AG10161, P30AG72975, R01AG15819, R01AG17917, R01AG036836, U01AG46152, U01AG61356, U01AG046139, P50 AG016574, R01 AG032990, U01AG046139, R01AG018023, U01AG006576, U01AG006786, R01AG025711, R01AG017216, R01AG003949, R01NS080820, U24NS072026, P30AG19610, U01AG046170, RF1AG057440, and U24AG061340, and the Cure PSP, Mayo and Michael J Fox foundations, Arizona Department of Health Services and the Arizona Biomedical Research Commission. We thank the participants of the Religious Order Study and Memory and Aging projects for the generous donation, the Sun Health Research Institute Brain and Body Donation Program, the Mayo Clinic Brain Bank, and the Mount Sinai/JJ Peters VA Medical Center NIH Brain and Tissue Repository. Data and analysis contributing investigators include Nilüfer Ertekin-Taner, Steven Younkin (Mayo Clinic, Jacksonville, FL), Todd Golde (University of Florida), Nathan Price (Institute for Systems Biology), David Bennett, Christopher Gaiteri (Rush University), Philip De Jager (Columbia University), Bin Zhang, Eric Schadt, Michelle Ehrlich, Vahram Haroutunian, Sam Gandy (Icahn School of Medicine at Mount Sinai), Koichi Iijima (National Center for Geriatrics and Gerontology, Japan), Scott Noggle (New York Stem Cell Foundation), Lara Mangravite (Sage Bionetworks). This study was supported by National Natural Science Foundation of China (T2225015, 62433008, 32200537, 31600826, 32371144), Shanghai Science and Technology Commission Program (23JS1410100), Key Science and Technology Project of Hainan Province (ZDYF2024SHFZ058), National Key Research and Development Program of China (2023YFF1204800, 2020YFA0712403, 2019YFA0111000), Lingang Laboratory & National Key Laboratory of Human Factors Engineering Joint Grant (LG-TKN-202203-01), Lingang Laboratory (LG-GG-202401-ADA010100, LG-GG-202401-ADA050100, LG-QS-202203-08), Shanghai Municipal Science and Technology Major Project (20JC1419500), Natural Science Foundation of Shanghai (20ZR1403100), the Hainan Academician Innovation Platform Fund, and the Hainan Province Clinical Medical Center. The funders had no role in the study design, data collection, data analysis, interpretation, or writing of this report.

CONFLICT OF INTEREST STATEMENT

The authors report no competing interests. Author disclosures are available in the Supporting Information.

CONSENT STATEMENT

The procedures of this study were approved by the Ethics Committee of Shanghai Sixth People's Hospital (approval number: 2019-032). All participants or their legal guardians provided written consent for research projects.

REFERENCES

1. Alzheimer's Association. 2019 Alzheimer's disease facts and figures. *Alzheimers Dement*. 2019;15(3):321-387. doi:10.1016/j.jalz.2019.01.010
2. Gatz M, Reynolds CA, Fratiglioni L, et al. Role of genes and environments for explaining Alzheimer disease. *Arch Gen Psychiatry*. 2006;63(2):168-174.
3. Bellenguez C, Küçükali F, Jansen IE, et al. New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nat Genet*. 2022;54(4):412-436.
4. Andrews SJ, Renton AE, Fulton-Howard B, Podlesny-Drabiniok A, Marcora E, Goate AM. The complex genetic architecture of Alzheimer's disease: novel insights and future directions. *EBioMed*. 2023;90:104511.
5. Sims R, Van Der Lee SJ, Naj AC, et al. Rare coding variants in PLCG2, ABI3, and TREM2 implicate microglial-mediated innate immunity in Alzheimer's disease. *Nat Genet*. 2017;49(9):1373-1384.
6. Holstege H, Hulsman M, Charbonnier C, et al. Exome sequencing identifies rare damaging variants in ATP8B4 and ABCA1 as risk factors for Alzheimer's disease. *Nat Genet*. 2022;54(12):1786-1794.
7. Logue MW, Schu M, Vardarajan BN, et al. Two rare AKAP9 variants are associated with Alzheimer's disease in African Americans. *Alzheimers Dement*. 2014;10(6):609-618.
8. Hartl D, May P, Gu W, et al. A rare loss-of-function variant of ADAM17 is associated with late-onset familial Alzheimer disease. *Mol Psychiatry*. 2020;25(3):629-639.
9. Shigemizu D, Asanomi Y, Akiyama S, Mitsumori R, Niida S, Ozaki K. Whole-genome sequencing reveals novel ethnicity-specific rare variants associated with Alzheimer's disease. *Mol Psychiatry*. 2022;27(5):2554-2562.
10. Cochran JN, Geier EG, Bonham LW, et al. Non-coding and loss-of-function coding variants in TET2 are associated with multiple neurodegenerative diseases. *Am J Hum Genet*. 2020;106(5):632-645.
11. Lambert J-C, Amouyel P. Genetic heterogeneity of Alzheimer's disease: complexity and advances. *Psychoneuroendocrinology*. 2007;32:S62-S70.
12. Yu J-T, Jiang T, Wang Y-L, et al. Triggering receptor expressed on myeloid cells 2 variant is rare in late-onset Alzheimer's disease in Han Chinese individuals. *Neurobiol Aging*. 2014;35(4):937.
13. Miyashita A, Wen Y, Kitamura N, et al. Lack of genetic association between TREM2 and late-onset Alzheimer's disease in a Japanese population. *J Alzheimers Dis*. 2014;41(4):1031-1038.
14. Wang B, Bao S, Zhang Z, et al. A rare variant in MLKL confers susceptibility to ApoE ε4-negative Alzheimer's disease in Hong Kong Chinese population. *Neurobiol Aging*. 2018;68:160.
15. Zhang D-F, Fan Y, Xu M, et al. Complement C7 is a novel risk gene for Alzheimer's disease in Han Chinese. *Natl Sci Rev*. 2019;6(2):257-274.
16. Qin W, Zhou A, Zuo X, et al. Exome sequencing revealed PDE11A as a novel candidate gene for early-onset Alzheimer's disease. *Hum Mol Genet*. 2021;30(9):811-822.
17. Luo R, Fan Y, Yang J, et al. A novel missense variant in ACAA1 contributes to early-onset Alzheimer's disease, impairs lysosomal

function, and facilitates amyloid-β pathology and cognitive decline. *Signal Transduct Target Ther*. 2021;6(1):1-16.
18. Kucukali F, Neumann A, Van Dongen J, et al. Whole-exome rare-variant analysis of Alzheimer's disease and related biomarker traits. *Alzheimers Dement*. 2023;19(6):2317-2331.
19. Katzman R, Zhang M, Wang Z, et al. A Chinese version of the mini-mental state examination; impact of illiteracy in a Shanghai dementia survey. *J Clin Epidemiol*. 1988;41(10):971-978.
20. Huang L, Chen K-L, Lin B-Y, et al. Chinese version of Montreal cognitive assessment basic for discrimination among different severities of Alzheimer's disease. *Neuropsychiatr Dis Treat*. 2018;14:2133-2140.
21. Pan F-F, Wang Y, Huang L, Huang Y, Guo Q-H. Validation of the Chinese version of Addenbrooke's cognitive examination III for detecting mild cognitive impairment. *Aging Ment Health*. 2022;26(2):384-391.
22. Zhao Q, Guo Q, Liang X, et al. Auditory verbal learning test is superior to Rey-Osterrieth complex figure memory for predicting mild cognitive impairment to Alzheimer's disease. *Curr Alzheimer Res*. 2015;12(6):520-526.
23. Zhao Q, Guo Q, Hong Z. Clustering and switching during a semantic verbal fluency test contribute to differential diagnosis of cognitive impairment. *Neurosci Bull*. 2013;29:75-82.
24. Kaplan E, Goodglass H, Weintraub S. Boston naming test. *Lea & Febiger*. 1983. https://search.worldcat.org/zh-cn/title/10450471
25. Zhao Q, Guo Q, Li F, Zhou Y, Wang B, Hong Z. The shape trail test: application of a new variant of the trail making test. *PLoS One*. 2013;8(2):e57333.
26. McKhann GM, Knopman DS, Chertkow H, et al. The diagnosis of dementia 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*. 2011;7(3):263-269.
27. Bondi MW, Edmonds EC, Jak AJ, et al. Neuropsychological criteria for mild cognitive impairment improves diagnostic precision, biomarker associations, and progression rates. *J Alzheimers Dis*. 2014;42(1):275-289.
28. Jessen F, Amariglio RE, Van Boxtel M, et al. A conceptual framework for research on subjective cognitive decline in preclinical Alzheimer's disease. *Alzheimers Dement*. 2014;10(6):844-852.
29. Huang L, Chen K, Liu Z, Guo Q. A conceptual framework for research on cognitive impairment with no dementia in memory clinic. *Curr Alzheimer Res*. 2020;17(6):517-525.
30. Kendig KI, Baheti S, Bockol MA, et al. Sentieon DNASeq variant calling workflow demonstrates strong computational performance and accuracy. *Front Genet*. 2019;10:736.
31. McLaren W, Gil L, Hunt SE, et al. The ensembl variant effect predictor. *Genome Biol*. 2016;17(1):1-14.
32. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. 2020;581:434-443. doi:10.1038/s41586-020-2308-7
33. Liu X, Jian X, Boerwinkle E. dbNSFP: a lightweight database of human nonsynonymous SNPs and their functional predictions. *Hum Mutat*. 2011;32(8):894-899.
34. Liu X, Li C, Mou C, Dong Y, Tu Y. dbNSFP v4: a comprehensive database of transcript-specific functional predictions and annotations for human nonsynonymous and splice-site SNVs. *Genome Med*. 2020;12(1):1-8.
35. Itan Y, Shang L, Boisson B, et al. The human gene damage index as a gene-level approach to prioritizing exome variants. *Proc Natl Acad Sci*. 2015;112(44):13615-13620.
36. Landrum MJ, Lee JM, Benson M, et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res*. 2018;46(D1):D1062-D1067.
37. Lee S, Emond MJ, Bamshad MJ, et al. Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *Am Hum Genet*. 2012;91(2):224-237.
38. Lee S, Teslovich TM, Boehnke M, Lin X. General framework for meta-analysis of rare variants in sequencing association studies. *Am Hum Genet*. 2013;93(1):42-53.
39. Shigemizu D, Mitsumori R, Akiyama S, et al. Ethnic and trans-ethnic genome-wide association studies identify new loci influencing Japanese Alzheimer's disease risk. *Transl Psychiatry*. 2021;11(1):151.
40. Schwartzentruber J, Cooper S, Liu JZ, et al. Genome-wide meta-analysis, fine-mapping and integrative prioritization implicate new Alzheimer's disease risk genes. *Nat Genet*. 2021;53(3):392-402.
41. Itan Y, Zhang S-Y, Vogt G, et al. The human gene connectome as a map of short cuts for morbid allele discovery. *Proc Natl Acad Sci*. 2013;110(14):5558-5563.
42. Raudvere U, Kolberg L, Kuzmin I, et al. g: Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res*. 2019;47(W1):W191-W198.
43. Rovelet-Lecrux A, Charbonnier C, Wallon D, et al. De novo deleterious genetic variations target a biological network centered on Aβ peptide in early-onset Alzheimer disease. *Mol Psychiatry*. 2015;20(9):1046-1056.
44. Lanoiselée H-M, Nicolas G, Wallon D, et al. APP, PSEN1, and PSEN2 mutations in early-onset Alzheimer disease: a genetic screening study of familial and sporadic cases. *PLoS Med*. 2017;14(3):e1002270.
45. Gu X, Zhao M, Han X, Liu L. Presenilin-1 mutation is associated with a hippocampus defect in Alzheimer's disease: meta-analysis for neuroimaging research. *Clin Neurol Neurosurg*. 2020;191:105679.
46. Sollis E, Mosaku A, Abid A, et al. The NHGRI-EBI GWAS catalog: knowledgebase and deposition resource. *Nucleic Acids Res*. 2023;51(D1):D977-D985.
47. Maslinski C, Fogel W. Catabolism of histamine. *Histamine and Histamine Antagonists*. Springer; 1991:165-189.
48. Cumming P, Vincent SR. Inhibition of histamine-N-methyltransferase (HNMT) by fragments of 9-amino-1, 2, 3, 4-tetrahydroacridine (tacrine) and by β-carbolines. *Biochem Pharmacol*. 1992;44(5):989-992.
49. Heidari A, Tongsook C, Najafipour R, et al. Mutations in the histamine N-methyltransferase gene, HNMT, are associated with nonsyndromic autosomal recessive intellectual disability. *Hum Mol Genet*. 2015;24(20):5697-5710.
50. Palada V, Terzić J, Mazzulli J, et al. Histamine N-methyltransferase Thr105Ile polymorphism is associated with Parkinson's disease. *Neurobiol Aging*. 2012(4):836.
51. Shan L, Bossers K, Unmehopa U, Bao A-M, Swaab DF. Alterations in the histaminergic system in Alzheimer's disease: a postmortem study. *Neurobiol Aging*. 2012;33(11):2585-2598.
52. Zlomuzica A, Dere D, Binder S, Silva MADS, Huston JP, Dere E. Neuronal histamine and cognitive symptoms in Alzheimer's disease. *Neuropharmacology*. 2016;106:135-145.
53. Stenmark H, Vitale G, Ullrich O, Zerial M. Rabaptin-5 is a direct effector of the small GTPase Rab5 in endocytic membrane fusion. *Cell*. 1995;83(3):423-432.
54. Palasca O, Santos A, Stolte C, Gorodkin J, Jensen LJ. TISSUES 2.0: an integrative web resource on mammalian tissue expression. *Database*. 2018;2018:bay003.
55. Novikova G, Kapoor M, Tcw J, et al. Integration of Alzheimer's disease genetics and myeloid genomics identifies disease risk regulatory elements and genes. *Nat Commun*. 2021;12(1):1-14.
56. Escott-Price V, Sims R, Bannister C, et al. Common polygenic variation enhances risk prediction for Alzheimer's disease. *Brain*. 2015;138(12):3673-3684.
57. Escott-Price V, Myers AJ, Huentelman M, Hardy J. Polygenic risk score analysis of pathologically confirmed Alzheimer disease. *Ann Neurol*. 2017;82(2):311-314.
58. Sims R, Hill M, Williams J. The multiplex model of the genetics of Alzheimer's disease. *Nat Neurosci*. 2020;23(3):311-322.

59. De Leon M, Convit A, DeSanti S, et al. The hippocampus in aging and Alzheimer's disease. *Neuroimaging Clin N Am*. 1995;5(1):1-17.
60. Horínek D, Varjassyová A, Hort J. Magnetic resonance analysis of amygdalar volume in Alzheimer's disease. *Curr Opin Psychiatry*. 2007;20(3):273-277.
61. de Jong LW, van der Hiele K, Veer IM, et al. Strongly reduced volumes of putamen and thalamus in Alzheimer's disease: an MRI study. *Brain*. 2008;131(12):3277-3285.
62. De Jager PL, Ma Y, McCabe C, et al. A multi-omic atlas of the human frontal cortex for aging and Alzheimer's disease research. *Sci Data*. 2018;5(1):1-13.
63. Bennett DA, Buchman AS, Boyle PA, Barnes LL, Wilson RS, Schneider JA. Religious orders study and rush memory and aging project. *J Alzheimers Dis*. 2018;64(S1):S161-S189.
64. Allen M, Carrasquillo MM, Funk C, et al. Human whole genome genotype and transcriptome data for Alzheimer's and other neurodegenerative diseases. *Sci Data*. 2016;3(1):1-10.
65. Wang M, Beckmann ND, Roussos P, et al. The Mount Sinai cohort of large-scale genomic, transcriptomic and proteomic data in Alzheimer's disease. *Sci Data*. 2018;5(1):1-16.
66. Serrano-Pozo A, Frosch MP, Masliah E, Hyman BT. Neuropathological alterations in Alzheimer disease. *Cold Spring Harb Perspect Med*. 2011;1(1):a006189.
67. Cataldo AM, Barnett JL, Pieroni C, Nixon RA. Increased neuronal endocytosis and protease delivery to early endosomes in sporadic Alzheimer's disease: neuropathologic evidence for a mechanism of increased β-amyloidogenesis. *J Neurosci*. 1997;17(16):6142-6151.
68. Bucci C, Parton RG, Mather IH, et al. The small GTPase rab5 functions as a regulatory factor in the early endocytic pathway. *Cell*. 1992;70(5):715-728.
69. Cataldo AM, Peterhoff CM, Troncoso JC, Gomez-Isla T, Hyman BT, Nixon RA. Endocytic pathway abnormalities precede amyloid β deposition in sporadic Alzheimer's disease and down syndrome: differential effects of APOE genotype and presenilin mutations. *Am J Pathol*. 2000;157(1):277-286.
70. Kuperstein I, Broersen K, Benilova I, et al. Neurotoxicity of Alzheimer's disease Aβ peptides is induced by small changes in the Aβ42 to Aβ40 ratio. *EMBO J*. 2010;29(19):3408-3420.
71. Haass C, Kaether C, Thinakaran G, Sisodia S. Trafficking and proteolytic processing of APP. *Cold Spring Harb Perspect Med*. 2012;2(5):a006270.
72. Gao S, Casey AE, Sargeant TJ, Mäkinen V-P. Genetic variation within endolysosomal system is associated with late-onset Alzheimer's disease. *Brain*. 2018;141(9):2711-2720.
73. Zhan L, Li J, Jew B, Sul JH. Rare variants in the endocytic pathway are associated with Alzheimer's disease, its related phenotypes, and functional consequences. *PLos Genet*. 2021;17(9):e1009772.
74. Laifenfeld D, Patzek LJ, McPhie DL, et al. Rab5 mediates an amyloid precursor protein signaling pathway that leads to apoptosis. *J Neurosci*. 2007;27(27):7141-7153.
75. Kitano M, Nakaya M, Nakamura T, Nagata S, Matsuda M. Imaging of Rab5 activity identifies essential regulators for phagosome maturation. *Nature*. 2008;453(7192):241-245.
76. Ubelmann F, Burrinha T, Salavessa L, et al. Bin1 and CD $_2$AP polarise the endocytic generation of beta-amyloid. *EMBO Rep*. 2017;18(1):102-122.
77. Kanatsu K, Morohashi Y, Suzuki M, et al. Decreased CALM expression reduces Aβ42 to total Aβ ratio through clathrin-mediated endocytosis of γ-secretase. *Nat Commun*. 2014;5(1):1-12.
78. Andersen OM, Bøgh N, Landau AM, et al. A genetically modified minipig model for Alzheimer's disease with SORL1 haploinsufficiency. *Cell Rep Med*. 2022;3(9):100740.
79. Haas HL, Sergeeva OA, Selbach O. Histamine in the nervous system. *Physiol Rev*. 2008;88(3):1183-1241.
80. Martorana A, Esposito Z, Koch G. Beyond the cholinergic hypothesis: do current drugs work in Alzheimer's disease?. *CNS Neurosci Ther*. 2010;16(4):235-245.
81. Panula P, Rinne J, Kuokkanen K, et al. Neuronal histamine deficit in Alzheimer's disease. *Neuroscience*. 1997;82(4):993-997.
82. Flores-Clemente C, Nicolás-Vázquez MI, Mera Jiménez E, Hernández-Rodríguez M. Inhibition of astrocytic histamine n-methyltransferase as a possible target for the treatment of Alzheimer's disease. *Biomolecules*. 2021;11(10):1408.
83. Miyashita A, Kikuchi M, Hara N, Ikeuchi T. Genetics of Alzheimer's disease: an East Asian perspective. *J Hum Genet*. 2023;68(3):115-124.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Cao J, Zhang C, Lo C-YZ, et al. Integrating rare pathogenic variant prioritization with gene-based association analysis to identify novel genes and relevant multimodal traits for Alzheimer's disease. *Alzheimer's Dement*. 2024;1-15. https://doi.org/10.1002/alz.14444

APPENDIX

MEMBERS OF THE ZIB CONSORTIUM

Jianfeng Feng[1,3,4], Xing-Ming Zhao[1,3,4,7,9], Gunter Schumann[1], Tianye Jia[1], Qihao Guo[5], Chun-Yi Zac Lo[19], Shuqiao Yao[10], Xiang Wang[10], Tianhong Zhang[11], Shenxun Shi[12], Qiang Luo[1], Jijun Wang[11], Jie Zhang[1], Xin Wang[13], Jing Ding[1], Dezhi Liu[14], Bo Yu[15], He Wang[1], Fei Li[16], Miao Cao[1], Chunshui Yu[17], Guang Yang[18], Xiao-Yong Zhang[1], Deniz Vatansever[1], Jingqi Chen[1,3,4]

[10]Medical Psychological Center, The Second Xiangya Hospital of Central South University, Changsha, China

[11]Shanghai Key Laboratory of Psychotic Disorders, Shanghai Mental Health Center, Shanghai Jiao Tong University School of Medicine, Shanghai, China

[12]Psychiatry Department, Fudan University Affiliated Huashan Hospital, Shanghai, China

[13]Department of Neurology, Zhongshan Hospital, Fudan University, Shanghai, China

[14]Department of Neurology, Shuguang Hospital Affiliated to Shanghai University of Traditional Chinese Medicine, Shanghai, China

[15]Department of Vascular Surgery, Shanghai Pudong Hospital, Fudan University Pudong Medical Center, Shanghai, China

[16]Ministry of Education and Shanghai Key Laboratory of Children's Environmental Health, Xinhua Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

[17]Department of Radiology and Tianjin Key Laboratory of Functional Imaging, Tianjin Medical University General Hospital, Tianjin, China

[18]Affiliation Department of Physics, Shanghai Key Laboratory of Magnetic Resonance, East China Normal University, Shanghai, China

[19]Department of Biomedical Engineering, Chung Yuan Christian University, Taoyuan, Taiwan, China