Genetics and Genomics
Epidemiology and Global Health

Reviewed Preprint
v2 • September 4, 2024
Revised by authors

Reviewed Preprint
v1 • July 1, 2024

Systematic evaluation of multifactorial causal associations for Alzheimer's disease and an interactive platform based on Mendelian randomization analysis —— MRAD

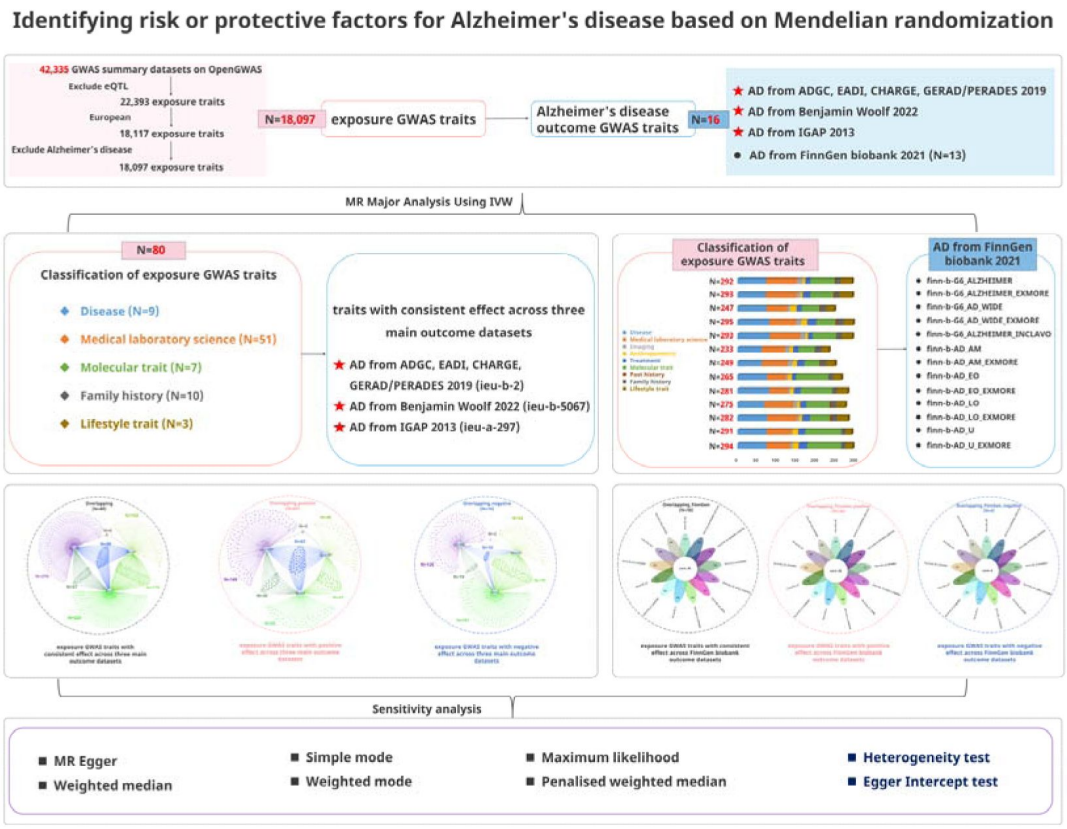
Tianyu Zhao, Hui Li, Meishuang Zhang, Yang Xu, Ming Zhang, Li Chen

Department of Pharmacology, College of Basic Medical Sciences, Jilin University, Changchun, China • Department of Neurology, Xuanwu Hospital, Capital Medical University, Beijing, China • Neurology and Intracranial Hypertension & Cerebral Venous Disease Center, National Health Commission of China, Xuanwu Hospital, Capital Medical University, Beijing, China • School of Nursing, Jilin University, Changchun, China

Abstract

Alzheimer's disease (AD) is a complex degenerative disease of the central nervous system. Traditional epidemiological studies have reported several risk factors for AD. However, most epidemiological studies are insufficient to draw definitive conclusions on causal association due to the potential for reverse causality and confounding bias. Therefore, elucidating its pathogenesis remains challenging. Mendelian randomization (MR) was developed for assessing causality using genetic variants as a new approach in epidemiological research. In this study, we used MR analysis to investigate potential AD risk factors to support extensive AD research. We used the inverse-variance weighted (IVW) model as the major analysis method to perform hypothesis-free Mendelian randomization analysis on the data from MRC IEU OpenGWAS (18,097 exposure traits and 16 AD outcome traits), and conducted sensitivity analysis with six models, to assess the robustness of the IVW results, to identify various classes of risk or protective factors for AD, early-onset AD, and late-onset AD. We generated 400,274 data entries in total, among which the major analysis method of IVW model consists of 73,129 records with 4840 exposure traits, which fall into 10 categories: Disease (n=17,168), Medical laboratory science (n=15,416), Imaging (n=4,896), Anthropometric (n=4,478), Treatment (n=4,546), Molecular trait (n=17,757), Gut microbiota (n=48), Past history (n=668), Family history (n=1,114), and Lifestyle trait (n=7,038). For the convenience of display and operation, an online platform called MRAD has been developed using the Shiny package with MR analysis results. MRAD can be freely accessed online at https://gwasmrad.com/mrad/. Moreover, novel potential AD therapeutic targets (CD33, TBCA, VPS29, GNAI3, PSME1) are identified, among which CD33 was positively associated with the main outcome traits of AD, as well as with both EOAD and LOAD. TBCA and VPS29 were negatively associated with the main outcome traits of AD, as well as with both EOAD and LOAD. GNAI3 and PSME1 were negatively associated with the main outcome traits of AD, as well as with LOAD, but had no significant causal association with EOAD. This is one of the most comprehensive studies in this field. The findings of our research advance understanding of the etiology of AD.

Graphical Abstract



eLife assessment

This **important** study introduces the MRAD database, an advancement in Alzheimer's disease research that provides a powerful tool for evaluating risk and protective factors through Mendelian randomization analysis. The evidence supporting the database's utility is **solid**, with findings backed by robust data, though addressing methodological concerns and ensuring more rigorous validation of associations would further strengthen its impact. This resource represents a significant leap forward in the field, offering unprecedented opportunities for researchers and clinicians to uncover key insights into Alzheimer's etiology, potentially revolutionizing how Alzheimer's research is approached and accelerating the discovery of new prevention strategies and treatments.

https://doi.org/10.7554/eLife.96224.2.sa3

Background

Alzheimer’s disease (AD) is a progressive degenerative disease of the central nervous system, characterized by cognitive impairment, reduced functional capacity for daily living, and behavioral changes. It can be divided into two types: early-onset AD (EOAD, age of onset ≤ 65 years) and late-onset AD (LOAD, age of onset > 65 years); the proportion of LOAD in patients with

AD is approximately 95%, with LOAD having a stronger genetic predisposition than EOAD[1–3]. According to the latest data from the World Health Organization (WHO), the population with AD is currently over 50 million worldwide and is expected to rise to 115 million by 2050[4,5]. With the increasing aging population, the incidence of AD continues to rise, making AD the fifth leading cause of death worldwide. Given that AD is a chronic complex disorder involving multiple pathophysiological changes, it is likely caused by the joint action of various factors in a multifaceted pathological process, and this intricate nature of AD contributes to the current challenges in its diagnosis and treatment, such as low consultation rates, high rates of misdiagnosis at initial consultations, and low rates of long-term standardized treatment[6], thereby making AD one of the most perplexing diseases. Consequently, examining the pathogenic mechanisms of AD, identifying its risk factors, and conducting timely and effective early screening and diagnosis are of utmost importance.

Traditional epidemiological studies have reported common risk factors for AD. Some metabolic co-morbidities are highly associated with AD, such as cardiovascular disease[7,8], obesity[9,10], and diabetes[11,12]. Serological parameters such as C-reactive protein[13], lipids[14,15], and vitamin levels[16–18] have been previously reported as potential biomarkers for AD. In addition, some factors related to lifestyle, family history, education, economic level, and environment correlate with AD[19–22]. However, most epidemiological studies are insufficient to draw definitive conclusions on causal association due to the potential for reverse causality and confounding bias.

Mendelian randomization (MR) analysis is an emerging method to explore the causal association between AD and various factors[23–25]. MR analysis reduces confounding and reverse causality due to the segregation and independent assortment of genes passed from parents to offspring[26]. In the absence of pleiotropy (that is, genetic variation related to a disease via other pathways) and demographic stratification, MR can present a clear estimate of risk of disease[27,28]. MR analysis is increasingly used to determine a causal relationship between potentially modifiable risk factors and outcomes[29]. These advantages make MR a valuable tool to better elucidate the potential risk or protective factors for AD.

Chen et al. [30] used MR analysis to reveal the causal relationship between AD and factors including sociodemographic and early life status. However, the study revealed they are restricted by the available variables from the UKB database, which lead to variables such as air pollution, blood glucose measures and so on were not included. And also, due to the high degree of heterogeneity present in AD subtypes, which have different biological and genetic characteristics. Thus, the previous studies cannot offer a systematic and complete viewpoint. Our study uses the MRC IEU OpenGWAS database as the sample source for MR analysis to address the aforementioned limitations. The MRC IEU OpenGWAS database, the largest open GWAS database globally, has compiled 42,335 GWAS summary datasets from sources such as the UK Biobank, FinnGen Biobank, and Biobank Japan. Analyzing large-scale datasets will break new ground for MR research on AD.

MR requires a combination of background knowledge in biology, computer science, software studies, and statistics, which often leads to a dilemma where biologists are not well-versed in computer and statistical fields, while computer science experts struggle to adopt a medical biology mindset. Consequently, the vast majority of available GWAS data have not been effectively utilized through MR. Therefore, the construction of a multi-level data platform specifically for AD based on MR analysis of massive GWAS data is of great strategic significance, and it will facilitate researchers and clinicians worldwide to conveniently and rapidly obtain risk factors that are causally associated with AD.

In summary, in this work we attempt to identify risk or protective factors causally associated with AD from a holistic and systematic perspective, thereby providing new ideas for understanding the AD pathogenesis, achieving early diagnosis, and developing clinical drugs. In the first place, this study uses a hypothesis free data mining approach to studying the possible etiology of Alzheimer's disease based on Mendelian randomization (MR), with specific attention to different AD subtypes (EOAD and LOAD). Based on this, we developed an online open integrated platform, MRAD (Mendelian randomization for Alzheimer's disease, *https://gwasmrad.com/mrad/*). Moreover, the platform was further enriched by including related targets' information such as functions and pathways retrieved from the public database Uniprot. The platform is the first multi-dimensional, integrated, shared, and interactive comprehensive platform for AD MR research to date.

Methods

Database and software

The following databases and software packages were used in this study: MRC IEU OpenGWAS[31] (*https://gwas.mrcieu.ac.uk/*), UniProt[32] (*https://www.uniprot.org/*), EVenn[33] (*http://www.ehbio.com/test/venn/%23/*), R (version 4.1.2) software[34].

MR design for AD (Figure 1)

Data sources

Exposure traits

Inclusion criteria: datasets of the European population.

Exclusion criteria: (i) eQTL-related datasets; (ii) AD-related datasets.

In this study, the GWAS datasets selected were derived from 42,335 GWAS datasets in the public database (MRC IEU OpenGWAS, *https://gwas.mrcieu.ac.uk/*). Based on the above inclusion and exclusion criteria, 19,942 eQTL-related datasets were excluded first, leaving 22,393 GWAS datasets. Next, the datasets with the European population were selected, and 18,117 GWAS datasets were obtained. Finally, 20 AD-related datasets were excluded; 18,097 GWAS datasets were obtained at the end as the exposure traits of this study (See Table S1 for basic information).

Outcome traits

Inclusion criteria: (i) datasets of patients with AD with complete information and clear data sources; (ii) datasets of the European population.

Exclusion criteria: (i) Number of SNPs <1 million; (ii) datasets with unspecified sex; (iii) datasets with a family history of AD; (iv) datasets with dementia.

Based on the above criteria, 16 GWAS datasets of outcome traits were selected from the MRC IEU OpenGWAS database, comprising datasets of AD from Alzheimer Disease Genetics Consortium (ADGC), Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium (CHARGE), The European Alzheimer's Disease Initiative (EADI), and Genetic and Environmental Risk in AD/Defining Genetic, Polygenic and Environmental Risk for Alzheimer's Disease Consortium (GERAD/PERADES) 2019 (ieu-b-2); AD from Benjamin Woolf 2022 (ieu-b-5067); AD from International Genomics of Alzheimer's Project (IGAP) 2013 (ieu-a-297) as the datasets of main outcome traits for AD, as well as 13 datasets from FinnGen biobank 2021 corresponding to various AD subtypes, referred to as AD-finn subtypes. (as shown in **Figure 2**)

Figure 1.

Study design

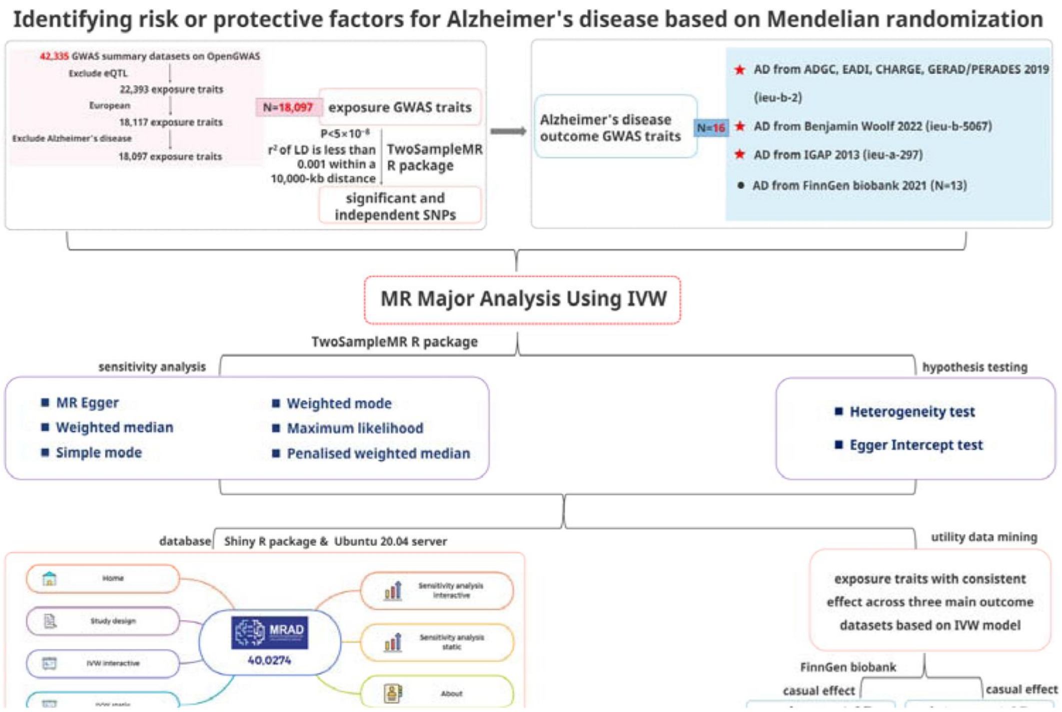


Figure 2.

Basic information of 16 outcome traits in MRC IEU OpenGWAS

GWAS ID	Trait	Consortium	Population	Sex	Year	Number of SNPs	Sample size	ncase	ncontrol	Category	Sub category	Author	PMID	Ontology	Unit	sd	Note
ieu-b-2	Alzheimer's disease	Alzheimer Disease Genetics Consortium (ADGC), European Alzheimer's Disease Initiative (EADI), Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium (CHARGE), Genetic and Environmental Risk in AD/Defining Genetic, Polygenic and Environmental Risk for Alzheimer's Disease Consortium (GERAD/PERADES),	European	Males and Females	2019	10528610	63926	21982	41944	Binary	Psychiatric / neurological	Kunkle BW	30820047	NA	NA	NA	NA
ieu-b-5067	Alzheimer's disease	NA	European	Males and Females	2022	12321875	488285	954	487331	NA	NA	Benjamin Woolf	NA	EFO:0006514:MONDO_0004975	NA	NA	GWAS created using BOLT-LMM. Pheno was algorithmically defined using the UKB. See https://www.mdpi.com/2072-6643/14/9/1697 for more details. Imputation panel: see https://doi.org/10.5523/bris.pnoat8cxo0u52p6ynfaekeiqi
ieu-a-297	Alzheimer's disease	IGAP	European	Males and Females	2013	7055882	54162	17008	37154	Disease	Psychiatric / neurological	Lambert	24162737	NA	log odds	NA	Effect allele frequencies are missing; forward(+) strand
finn-b-G6_ALZHEIMER	Alzheimer disease	NA	European	Males and Females	2021	16380466	NA	3899	214893	Binary	NA	NA	NA	NA	NA	NA	G6_ALZHEIMER
finn-b-G6_ALZHEIMER_EXMORE	Alzheimer disease (more controls excluded)	NA	European	Males and Females	2021	16379528	NA	3899	111471	Binary	NA	NA	NA	NA	NA	NA	G6_ALZHEIMER_EXMORE
finn-b-G6_AD_WIDE	Alzheimer<c3><95>s disease, wide definition	NA	European	Males and Females	2021	16380462	NA	5918	212874	Binary	NA	NA	NA	NA	NA	NA	G6_AD_WIDE
finn-b-G6_AD_WIDE_EXMORE	"Alzheimer<c3><95>s disease, wide definition" (more controls excluded)	NA	European	Males and Females	2021	16379561	NA	5918	111471	Binary	NA	NA	NA	NA	NA	NA	G6_AD_WIDE_EXMORE
finn-b-G6_ALZHEIMER_INCLAVO	Alzheimer disease, including avohilmo	NA	European	Males and Females	2021	16380466	NA	4166	214626	Binary	NA	NA	NA	NA	NA	NA	G6_ALZHEIMER_INCLAVO
finn-b-AD_AM	Alzheimer<c3><95>s disease (Atypical or mixed)	NA	European	Males and Females	2021	16380464	NA	800	214893	Binary	NA	NA	NA	NA	NA	NA	AD_AM
finn-b-AD_AM_EXMORE	Alzheimer<c3><95>s disease (Atypical or mixed) (more controls excluded)	NA	European	Males and Females	2021	16379412	NA	800	111471	Binary	NA	NA	NA	NA	NA	NA	AD_AM_EXMORE
finn-b-AD_EO	Alzheimer<c3><95>s disease (Early onset)	NA	European	Males and Females	2021	16380464	NA	587	214885	Binary	NA	NA	NA	NA	NA	NA	AD_EO
finn-b-AD_EO_EXMORE	Alzheimer<c3><95>s disease (Early onset) (more controls excluded)	NA	European	Males and Females	2021	16379428	NA	587	111471	Binary	NA	NA	NA	NA	NA	NA	AD_EO_EXMORE
finn-b-AD_LO	Alzheimer<c3><95>s disease (Late onset)	NA	European	Males and Females	2021	16380466	NA	2670	214871	Binary	NA	NA	NA	NA	NA	NA	AD_LO
finn-b-AD_LO_EXMORE	Alzheimer<c3><95>s disease (Late onset) (more controls excluded)	NA	European	Males and Females	2021	16379476	NA	2670	111471	Binary	NA	NA	NA	NA	NA	NA	AD_LO_EXMORE
finn-b-AD_U	Alzheimer<c3><95>s disease (undefined)	NA	European	Males and Females	2021	16380464	NA	159	214893	Binary	NA	NA	NA	NA	NA	NA	AD_U
finn-b-AD_U_EXMORE	Alzheimer<c3><95>s disease (undefined) (more controls excluded)	NA	European	Males and Females	2021	16379392	NA	159	111471	Binary	NA	NA	NA	NA	NA	NA	AD_U_EXMORE

Selection of instrumental variables

SNPs serve as instrumental variables for MR research. In this study, 18,097 exposures-variable SNPs were selected for MR research from the GWAS data (as mentioned in *Exposure traits*) respectively, with the selected SNPs fulfilling the following requirements: (i) a genome-wide significant association with risk factors ($p < 5\times10^{-8}$) in the European 1000 Genomes Project reference panel; (ii) independent of one another (that is, the r^2 of linkage disequilibrium (LD) is less than 0.001 within a 10,000-kb distance) to avoid potential biases caused by LD between SNPs in the analysis.

Statistical models for causal effect inference

A random-effects IVW model was used in this study as the major analysis method to uncover potential risk or protective factors for AD. The random-effects IVW model as the gold standard for MR studies, its principle is to calculate the inverse of the variance of each IV as its weight, assuming all IVs are valid. The regression does not include an intercept term, and the final result is the weighted average of the effect estimates from all IVs [35]. This model indicates that the true effect values may vary across different studies due to both sampling error and the heterogeneity of the true effect. The weight of each study is jointly determined by its inverse variance and the estimated heterogeneity variance. Thus, as long as there is no pleiotropy, even when there is significant heterogeneity ($p < 0.05$), this method remains the best MR model.

To assess the robustness of the IVW results, sensitivity analysis was performed using six additional models: (i) MR-Egger: MR-Egger's biggest difference from IVW is that it considers the intercept term during regression to evaluate bias caused by horizontal pleiotropy. The intercept represents the magnitude of horizontal pleiotropy, with a value close to 0 indicating minimal pleiotropy. The primary purpose is to detect and correct for horizontal pleiotropy. Thus, when significant horizontal pleiotropy is observed ($p < 0.05$), this method is preferred [36,37]. (ii) Weighted median: The weighted median method is a technique for evaluating causal relationships using a majority of genetic variants (SNPs). If at least 50% of the SNPs are valid IVs, the median of the causal estimates will tend toward the true causal effect. This method provides an unbiased estimate (i.e., the "majority validity" assumption) [38]. (iii) Simple mode: Involves comparing the frequencies or proportions of genotypes or phenotypes between control and experimental groups. Moreover, it can illustrate whether the observed differences in genotypes or phenotypes between the two groups are statistically significant. (iv) Weighted mode: The weighted mode method is a technique for combining multiple Mendelian randomization estimates. This method assigns weights to the causal effect estimates of different genetic variants on the trait and then takes the weighted mode as the final estimate of the causal effect. In genetic variant estimates, the method can decrease bias caused by outliers. (v) Maximum likelihood: This method is used when it is known that a random sample follows a particular probability distribution; however, the specific parameters of that distribution remain unknown, and it involves conducting multiple experiments, observing the results, and using those results to infer the approximate values of the parameters [39]. (vi) Penalized weighted median: An enhanced version of the weighted median estimate that provides a consistent estimate of the causal effect. (vii) Heterogeneity and horizontal pleiotropy assessment use the heterogeneity tests [40] and Egger intercept tests [41], respectively.

The above analyses were performed using the TwoSampleMR[42] package in the R (version 4.1.2) software. Association of exposures with outcomes was assessed using odds ratio (OR) and 95% confidence interval (95% CI), with $OR > 1$ indicating a positive association (risk factor) and $0 < OR < 1$ indicating a negative association (protective factor). Differences with a two-sided $p < .05$ were considered statistically significant. Furthermore, owing to the relatively large number of exposure

and outcome traits included in this study, the multiple testing correction method Bonferroni correction was added to identify significant hits, threshold for Bonferroni-corrected was 0.05 divided by 289,552 tests (p <1.727e-07).

Building the MRAD platform

In this study, the online MRAD platform was developed using the Shiny package[43] in R (version 4.1.2) and hosted on an Ubuntu 20.04 server. By leveraging Shiny, we combined the computational capabilities of R with modern web technologies, allowing to construct an interactive user interface with novel approaches.

Results

Results of hypothesis-free Mendelian randomization analysis for Alzheimer's disease

Based on hypothesis-free Mendelian randomization analysis for Alzheimer's disease, this study generated a total of 400,274 data points. The major analysis method of IVW model consists of 73,129 records with 4840 exposure traits, which fall into 10 categories: Disease (n=17,168), Medical laboratory science (n=15,416), Imaging (n=4,896), Anthropometric (n=4,478), Treatment (n=4,546), Molecular trait (n=17,757), Gut microbiota (n=48), Past history (n=668), Family history (n=1,114), and Lifestyle trait (n=7,038), as shown in **Figure 3**. To assess the robustness of the IVW results, sensitivity analysis was performed using six other models (MR-Egger with a total of 50,804 records, Weighted median with a total of 50,804 records, Simple mode with a total of 50,804 records, Weighted mode with a total of 50,804 records, Maximum likelihood with a total of 73,125 records, and Penalized weighted median with a total of 50,804 records).

MRAD platform integration

Based on the 400,274 data points stated above, we created herein is an online data analysis platform for identifying the risk or protective factors for AD called MRAD (Mendelian randomization for Alzheimer's disease, *https://gwasmrad.com/mrad/*). It contains six modules: (i) Home; (ii) Study Design; (iii) IVW interactive; (iv) IVW static; (v) Sensitivity analysis interactive; and (vi) Sensitivity analysis static; The platform provides a user-friendly search interface, allowing users to search, interactively visualize, analyze, and download the obtained results (MRAD User Guide see Supplementary Material for details). In our view, as the first interactive comprehensive platform for AD MR research to date, this online platform would benefit the field of scientific research in AD in numerous ways. On the one hand, it would allow researchers to quickly identify risk or protective factors from their own research and generate novel hypothesis regarding the molecular mechanism of AD. On the other hand, it would allow researchers with complementary expertise to provide multiple characterizations of the same data. As the platform is hosted on a server and accessed through a web interface, which could meet the multi-terminal compatibility, thereby MRAD's online presence could increase access to potential users.

MRAD utility data mining

To demonstrate the utility of MRAD platform, we focus on the IVW model-identified exposure traits that have significantly and consistently effect across three main outcome traits of AD to demonstrate the performance of the MRAD platform. Detailed investigation and reporting of other factors will be carried out in future research.

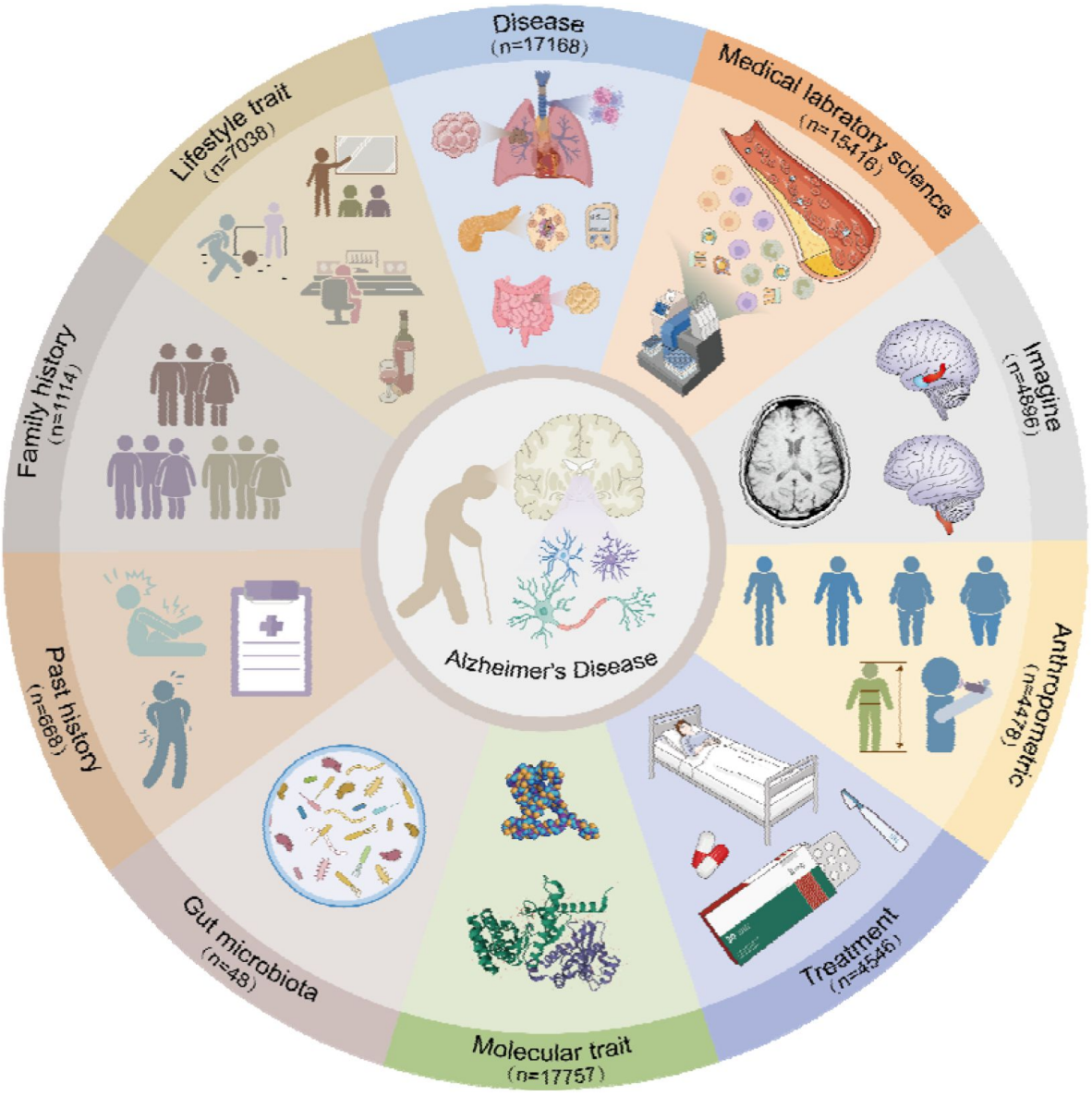


Figure 3.

Categories of the exposure traits identified by IVW model

In this study, MR analysis was first performed on the three main outcome traits of AD to explore their potential risk or protective factors, leading to identification of a total of 80 exposure traits ($p<0.05$), which fell into five Classification I categories: Medical laboratory science (n=51), Family history (n=10), Disease (n=9), Molecular trait (n=7), and Lifestyle trait (n=3). A total of 63 exposure traits (risk factors) were positively associated with all the three main outcome traits, while 16 exposure traits (protective factors) were negatively associated with the three main outcome traits, with Ulcerative colitis (ebi-a-GCST000964) being negatively associated with the AD outcome traits of ieu-b-2 and ieu-a-297, and positively associated with the AD outcome traits of ieu-b-5067. MR analysis was performed on the outcome traits of 13 different AD-finn subtypes to further examine the causal association between the above-identified key common exposure traits and different subtypes of AD outcome traits. The results are provided below in detail.

Causal association between medical laboratory science and the main outcome traits of AD

In this study, the 51 medical laboratory science items that each had a causal effect on the main outcome traits of AD were grouped into three Classification II categories (blood lipids and lipoproteins (n=36), immunological tests (n=12), and plasma protein tests (n=3)).

1 Blood lipids and lipoproteins

A total of 36 blood lipids and lipoproteins items as exposure traits had effects on the main outcome traits of AD: (1) 32 of which were positively associated with the main outcome traits, 7 of which, e.g., apolipoprotein B (ieu-b-108), were positively associated with EOAD (finn-b-AD_EO) and LOAD (finn-b-AD_LO); free cholesterol in IDL (met-c-868) was positively associated with EOAD (finn-b-AD_EO); 4 of which, e.g., phospholipids in small LDL (met-d-S_LDL_PL), were positively associated with LOAD (finn-b-AD_LO), as shown in **Figure 4A**. The corresponding sensitivity analysis and Bonferroni correction results are shown in Figure S1 and Table S2. (2) four of which were negatively associated with the main outcome traits, apolipoprotein A-I (ieu-b-107) was negatively associated with both EOAD (finn-b-AD_EO) and LOAD (finn-b-AD_LO), and the negative causal association was slightly stronger for EOAD than for LOAD; phospholipids to total lipids ratio in chylomicrons and extremely large VLDL (met-d-XXL_VLDL_PL_pct) was negatively associated with LOAD (finn-b-AD_LO). These findings are illustrated in **Figure 4B**. The corresponding sensitivity analysis and Bonferroni correction results are shown in Figure S2 and Table S2.

2 Immunological tests

A total of 12 immunological test items as exposure traits had positive effects on the main outcome traits of AD. Six of which, e.g., CD33 on Monocytic Myeloid-Derived Suppressor Cells (ebi-a-GCST90001952), were positively associated with LOAD (finn-b-AD_LO), as shown in **Figure 4C**. The corresponding sensitivity analysis and Bonferroni correction results are shown in Figure S3 and Table S2.

3 Plasma protein tests

A total of 3 plasma protein tests items as exposure traits had negative effects on the main outcome traits of AD. The three exposure traits were C-reactive protein (ukb-d-30710_raw, ukb-d-30710_irnt, and ieu-b-4764). All the three exposure traits were negatively associated with EOAD (finn-b-AD_EO) and LOAD (finn-b-AD_LO), as shown in **Figure 4D**. The corresponding sensitivity analysis and Bonferroni correction results are shown in Figure S4 and Table S2.

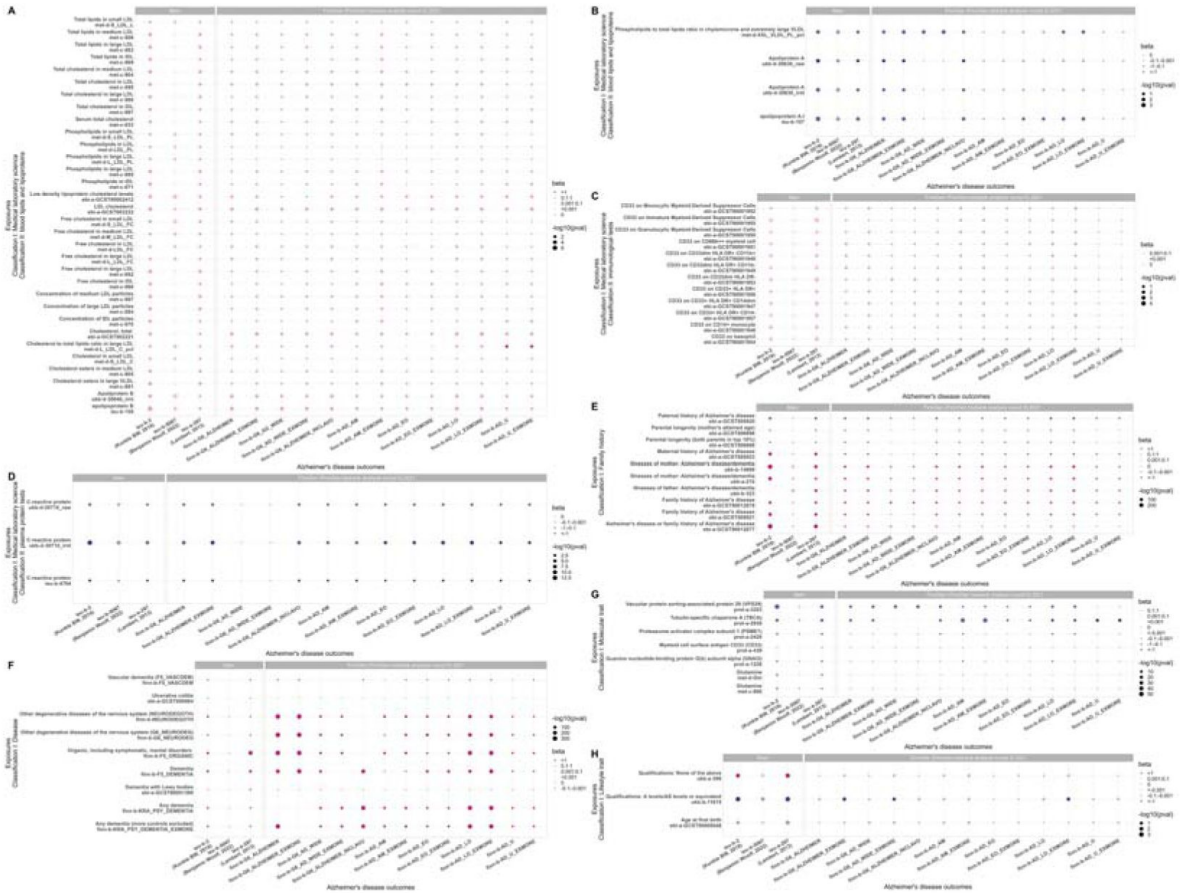


Figure 4.

80 exposure traits with causal effects on the main outcome traits of AD based on major analysis method random-effects IVW model.

Figure 4A. Thirty-two blood lipids and lipoproteins items that were positively associated with the main outcome traits of AD. **Figure 4B**. Four blood lipids and lipoproteins items that were negatively associated with the main outcome traits of AD. **Figure 4C**. Twelve immunological test items that were positively associated with the main outcome traits of AD. **Figure 4D**. Three plasma protein tests items that were negatively associated with the main outcome traits of AD. **Figure 4E**. Ten family history items with causal effects on the main outcome traits of AD. **Figure 4F**. Nine diseases items with causal effects on the main outcome traits of AD. **Figure 4G**. Seven molecular trait items with causal effects on the main outcome traits of AD. **Figure 4H**. Three lifestyle trait items with causal effects on the main outcome traits of AD.
Note: The pink dots in the figure represent positive association, the blue dots in the figure represent negative association, with the color depth of the dots being positively proportional to the OR value (the darker the color, the larger the OR value), and the size of the dots being inversely proportional to the p-value (the smaller the p-value, the larger the dots). The gray dots represent no significant causal association (p>0.05).

Causal association between family history and the main outcome traits of AD

A total of 10 family history items as exposure traits had causal effects on the main outcome traits of AD. In particular, a parental or family history of AD increased the overall risk of developing AD, and was positively associated with both EOAD (finn-b-AD_EO) and LOAD (finn-b-AD_LO), as shown in **Figure 4E**. The corresponding sensitivity analysis and Bonferroni correction results are shown in Figure S5 and Table S2.

Causal association between diseases and the main outcome traits of AD

In this study, the 9 diseases items that each had a causal effect on the main outcome traits of AD were grouped into four Classification II categories (dementia (n=5), neurodegenerative diseases (n=2), mental disorders associated with neurological diseases (n=1), and digestive system diseases (n=1)). Their causal effects with the main outcome traits of AD and the outcome traits of EOAD (finn-b-AD_EO) and LOAD (finn-b-AD_LO) are shown in **Figure 4F**. The corresponding sensitivity analysis and Bonferroni correction results are shown in Figure S6 and Table S2.

Causal association of molecular traits with the main outcome traits of AD

A total of 7 molecular trait items as exposure traits had causal effects on the main outcome traits of AD, among which Myeloid cell surface antigen CD33 (prot-a-439) was positively associated with the main outcome traits of AD, as well as with both EOAD (finn-b-AD_EO) and LOAD (finn-b-AD_LO). The remaining six were all negatively associated with the main outcome traits of AD, and their causal effects on the outcome traits of 13 AD-finn subtypes were as follows: (i) tubulin-specific chaperone A (TBCA; prot-a-2930) and vacuolar protein sorting-associated protein 29 (VPS29; prot-a-3203) were negatively associated with both EOAD (finn-b-AD_EO) and LOAD (finn-b-AD_LO); (ii) guanine nucleotide-binding protein G(k) subunit alpha (GNAI3; prot-a-1226) and proteasome activator complex subunit 1 (PSME1; prot-a-2420) were negatively associated with LOAD (finn-b-AD_LO), but had no significant causal association with EOAD (finn-b-AD_EO) ($p>0.05$); and (iii) neither glutamine (met-c-860) nor glutamine (met-d-Gln) had significant causal association with EOAD (finn-b-AD_EO) or LOAD (finn-b-AD_EO) ($p>0.05$), as shown in **Figure 4G**. The corresponding sensitivity analysis and Bonferroni correction results are shown in **Figure 5** and Table S2.

Causal association of lifestyle traits with the main outcome traits of AD

A total of 3 lifestyle trait items as exposure traits had causal effects on the main outcome traits of AD. Their causal effects with the main outcome traits of AD and the outcome traits of EOAD (finn-b-AD_EO) and LOAD (finn-b-AD_LO) are shown in Figure4H. The corresponding sensitivity analysis and Bonferroni correction results are shown in Figure S7 and Table S2.

Discussion

Despite decades of research on AD, controversy still remains regarding which factors play an important in its pathogenesis. This study carried out hypothesis-free Mendelian randomization analysis for Alzheimer's disease, which provided a thorough and comprehensive evaluation with regard to risk or protective factors for AD. This MR study covers most exposure traits that are causally associated with AD outcome traits, including diseases, medical laboratory science items, imaging items, anthropometric items, treatments, molecular traits, gut microbiota, past histories, family histories, and lifestyle traits, and reveals the causal associations between these exposure traits and different AD subtypes.

Figure 5.

Statistical models for causal effect results of seven molecular trait items with causal effects on the main outcome traits of AD.

Note:
(i) For column Inverse variance weighted, MR egger, Weighted median, Simple mode, Weighted mode, Maximum likelihood, and Penalized weighted median: the pink dots in the figure represent positive association, the blue dots represent negative association, with the color depth of the dots being positively proportional to the OR value (the darker the color, the larger the OR value), and the size of the dots being inversely proportional to the p-value (the smaller the p-value, the larger the dots). The gray dots represent no significant causal association (p>0.05). The star mark(✪) represents that is significant at the Bonferroni threshold (p<1.727e-07).
(ii) For column Heterogeneity test: the pink dots in the figure represent the effect of heterogeneity was considered negligible (heterogeneity_pval> 0.05).
The gray dots represent significant association (p<0.05).
(iii) For column Egger intercept test: the pink dots in the figure represent there was no significant difference between Egger Intercept and 0, indicating no horizontal pleiotropy (Horizontal_pval> 0.05). The gray dots represent significant association (p<0.05). The dark gray dots represent not applicable due to the quantity of SNP was less than 3.

Based on this, for the convenience of display and operation, a user-friendly prediction platform was built online called MRAD. The MRAD provides a one-stop online analysis service for researchers worldwide, including data retrieval → visualization → personalized analysis → data download. Users can obtain analysis results of different MR models (the main IVW model and six sensitivity analysis models) on 18,097 exposure traits and 16 AD outcome traits, totaling 400,274 records, and are allowed to set personalized parameters to meet different analysis needs. Additionally, the MRAD provides interactive visualization interfaces and download functions for the above results.

MRAD platform provides a unique resource for systematically identifying risk or protective factors of AD, which facilitates early identification, diagnosis, prevention, and treatment, with significant clinical and social value. It could have several strengths: (i) The current methods for identifying AD mainly rely on assessment scales, cerebrospinal fluid (CSF) examinations, and brain PET/MRI. However, assessment scales can be biased by factors such as the anxiety and nervousness of the subjects. CSF examinations require an invasive lumbar puncture, leading to low patient acceptance. PET/MRI scans are expensive and have limited equipment accessibility. These limitations restrict early AD identification. Thus, there is a pressing clinical need for readily available, time- and cost-effective, and accurate detection methods. In this study, the Medical laboratory science and Molecular trait used could be less expensive, faster to detect, easier to operate, and more accessible for widespread adoption. They hold great value for early AD identification and have the potential to become crucial tools for identifying AD in the future. (ii) Imaging acts as a powerful assistive tool for diagnosing Alzheimer's disease. Traditional imaging examinations mainly depict changes in the brain's macroscopic structure, while research on microstructural changes in disease-related areas is relatively limited. Studies have demonstrated that microstructural neurodegenerative processes are extensive and pronounced during AD progression. Our study results cover traditional macroscopic neuroimaging results and reveal numerous potential causal relationships between brain microstructure and AD. The combination of macroscopic and microstructural insights will provide more valuable information for clinical diagnosis. (iii) Clarifying patient's disease, past history, and family history can aid in preventing AD at an early stage, and prevention of AD could be attained through monitoring anthropometric indicators, improving gut microbiota, and adjusting lifestyle traits. (iv) Currently, the development of new drugs for AD is mainly underscored by Aβ, Tau, and other inhibitors. Since 2000, global pharmaceutical companies have invested hundreds of billions of dollars in the development of new drugs for AD, and these drugs have not yielded successful results. AD drug development has thus been perceived as having the highest failure rate of all drug research, reaching 99.6%. Hence, further research on molecular traits to find new targets and develop new drugs for these targets will provide new pathways for AD treatment.

To briefly demonstrate the performance of MRAD, we explored the IVW model-identified exposure traits that had significantly consistently effect across all the three main outcome traits of AD.

The association of lipids and lipoproteins, C-reactive protein, family histories, neurological disorders, glutamine, and education level with AD has been widely reported[23,44,67] and is consistent with the results of this study. Moreover, given that the prevalence of LOAD is about 95% in patients with AD and that LOAD has a stronger genetic predisposition than EOAD[1–3], identifying new risk genes for LOAD is crucial for understanding its potential etiology. Therefore, this study further explored the relationships between these traits and different AD subtypes, leading to the following findings: (i) apolipoprotein B, cholesterol, total, LDL cholesterol, Low density lipoprotein cholesterol levels, total cholesterol in LDL, total cholesterol in medium LDL, cholesterol to total lipids ratio in large LDL, free cholesterol in large LDL, free cholesterol in LDL, phospholipids in small LDL, parental or family history of AD, parental longevity (mother's attained age), dementia, vascular dementia, dementia with Lewy bodies, other degenerative diseases of the nervous system, and organic, including symptomatic, mental disorders were all positively associated with LOAD; (ii) apolipoprotein A-I, phospholipids to total lipids ratio in chylomicrons

and extremely large VLDL, C-reactive protein, parental longevity (both parents in top 10%), and qualifications: A levels/AS levels or equivalent were all negatively associated with LOAD. These findings suggest that the above traits may have critical impacts on LOAD.

Moreover, some novel potential therapeutic targets of AD were identified as follows: CD33 on Monocytic Myeloid-Derived Suppressor Cells, CD33 on CD33+ HLA DR+ CD14dim, CD33 on CD33+ HLA DR+, CD33 on CD33+ HLA DR+ CD14-, CD33 on CD33dim HLA DR -, CD33 on CD33dim HLA DR+ CD11b-, and Myeloid cell surface antigen CD33 were positively associated with all the three main outcome traits of AD and the risk of LOAD. It has been reported that CD33 is a 67 kDa glycosylated transmembrane protein, a member of the sialic acid-binding immunoglobulin like lectins family (SIGLECS family), which is an important receptor for cell growth and survival, as well as a critical receptor for the clathrin-independent endocytosis pathway and the innate and adaptive immune system functions. CD33 is mainly expressed in microglia, which are a type of glial cells in the central nervous system[68]. Meanwhile, the splicing efficiency of CD33 affects microglia activation[69]. Several genome-wide association studies have demonstrated that CD33 is a high-risk gene for AD[70–71]. In animal models, knockdown of CD33 significantly reduced amyloid plaque levels and knockout mice did not exhibit other health defects. Sialylated glycoproteins and glycolipids on amyloid plaques bind to CD33, which is most likely the cause of the amyloid "immune escape"[72]. Furthermore, polymorphisms in CD33 can increase the risk of AD by causing neuronal degeneration in the hippocampal and parahippocampal regions of the brain[73]. Downregulation of the sialic acid-binding domain of CD33 can reduce the risk of developing AD. Therefore, inhibiting CD33 is an effective approach to inhibit the development of AD, and the sialic acid-binding site on CD33 is a promising pharmacophore[74].

Tubulin-specific chaperone A (TBCA) was negatively associated with all the three main outcome traits of AD, as well as EOAD and LOAD (pval is significant at the Bonferroni threshold). TBCA is an important member of the tubulin-specific chaperones (TBCs) family. Tian et al. and Nolasco et al. demonstrated that TBCA can regulate the proportion of α and β-tubulin, enabling them to correctly aggregate into cellular microtubules[75]. Cellular microtubules play important roles in many biological functions, especially in cell movement, cell division, intracellular transport, and cell structure. After silencing TBCA, abnormal microtubule aggregation occurs in mammalian cells, and the cells cannot grow and divide normally, ultimately leading to apoptosis[76,77]. Moreover, studies have shown that TBCA plays a crucial role in correct β-tubulin folding and α/β-tubulin heterodimer formation[78]. Protein misfolding can lead to many diseases, such as neurodegenerative diseases. Additionally, higher levels of TBCA are significantly associated with lower AD risk[79]. These findings suggest that TBCA may serve as a potential protective factor against AD.

Vacuolar protein sorting-associated protein 29 (VPS29) was negatively associated with all the three main outcome traits of AD, as well as EOAD and LOAD (pval is significant at the Bonferroni threshold). VPS29 is a component of the retromer complex and is highly expressed in the brain, heart, and kidneys, playing an essential role in retromer functions such as synaptic transmission, survival, and movement[80]. Retromer mainly consists of the VPS26-VPS29-VPS35 trimer and Sorting Nexins (SNXs), and its defects are closely related to various human diseases, including neurodegenerative diseases[80]. Studies have reported that VPS29 knockdown leads to reduced levels of VPS35 and VPS26[81,82], which regulates the localization of retromer within neurons and is essential for the aging nervous system[80]. The retromer complex has been found to regulate the transport of a variety of substances, including amyloid precursor protein (APP), β-secretase, and phagocytic receptors on microglia. The retromer complex regulates the production of amyloid-β (Aβ) by regulating the transport of relevant carrier proteins, thus playing a role in AD[83]. When the retromer complex malfunctions, the pathway for the reverse transport of APP and β-secretase to the trans-Golgi network is disrupted, resulting in an increase in the production of Aβ, which accelerates the pathological process of AD[84]. Meanwhile, the reduction of phagocytic receptors on the surface of microglia weakens the clearance and protective functions

of microglia. Recent studies have shown that stabilizing the retromer complex through chaperone proteins can limit the amyloid processing of APP to reduce the production of Aβ[83]. These findings suggest that the retromer complex can serve as a new therapeutic target to intervene in the pathological progression of AD.

Guanine nucleotide-binding protein G(k) subunit alpha (GNAI3) was negatively associated with the three main outcome traits of AD and the risk of LOAD. G proteins are a class of signal transduction proteins that can bind with guanosine diphosphate (GDP) and have guanosine triphosphate (GTP) hydrolysis activity; they have more than 40 types, consisting of alpha, beta, and gamma subunits with a total molecular weight of about 100 kDa, with the alpha subunit having the greatest variation and determining the specificity of the G proteins[85]. G proteins are intracellular membrane proteins that shuttle between receptors and effector proteins, acting as signal transducers and playing an absolute dominant role in transmembrane cell signaling in the body. All cellular activities are related to signals, and signals are the initiating factors of all cell activities, while physiological responses are only the final results of signals acting on cells. After receiving external stimuli, cells respond by implementing signal transduction through a set of specific mechanisms to ultimately regulate the expression of specific genes, and the whole process is referred to as a cellular signaling pathway. In the pathogenesis of AD, the abnormal content and distribution of multiple signaling molecules, as well as the abnormality of signa transmission pathways, play an important role in AD pathological changes[86], suggesting that gaining insights into signal transduction mechanisms may provide a potential new pathway to explore the pathogenesis of AD.

Proteasome activator complex subunit 1 (PSME1) was negatively associated with all the three main outcome traits of AD and the risk of LOAD. PSME1 is the encoding gene of the 11s proteasome activator subunit (also known as PA28α) and is located on human chromosome 14q11.2. PA28α is an activator of proteasome, which mainly increases the protein degradation activity of 20S proteasome and participates in MHC-I (major histocompatibility complex I) restricted antigen presentation[87]. Studies have shown that PA28α overexpression in the brain of female mice can effectively prevent protein aggregation in the hippocampus, thereby reducing depression-like behavior and enhancing learning and memory ability[88]. Related studies have shown that proteasome function and PA28α expression are inhibited in the brain of diabetic rats[88]. The PA28 expression in the diabetic brain has a certain regulatory effect on protein metabolism caused by oxidative damage[88]. As suggested above, PSME1 may be a new potential therapeutic target for AD and deserves further investigation.

Conclusions

To the best of our knowledge, this is one of the most comprehensive studies to provide important insight into genetic etiology underlying AD based on hypothesis-free Mendelian randomization analysis. In the meantime, we developed the first MR platform for AD, of great clinical and scientific significance that provided a thorough and comprehensive evaluation with regard to risk or protective factors for AD. It also provided physicians and scientists with a very convenient, free as well as user-friendly tool for further scientific investigation. It is important to notice that we recognized CD33, TBCA, VPS29, GNAI3, and PSME1 as novel potential therapeutic targets for AD that deserve further investigation in more detail. However, in this study, since the GWAS datasets for both the exposure and the outcome traits (AD) selected were obtained from the public database (MRC IEU OpenGWAS), where the GWAS datasets for AD are only of European population, and since we use the TwoSampleMR, which requires that the populations for the exposure traits and the outcome traits be the same to satisfy the requirement for a control variable, this study currently has certain limitations in terms of population. We initiated a Mendelian randomization study on AD at clinical hospitals in China and are currently in the sample collection stage to

address the limitations. In the future, we will integrate data from more populations and continuously update new advances in AD research to explore its potential differences in different populations.

Declarations

Ethics approval and consent to participate

Not applicable. All data in this study are sourced from publicly available datasets.

Consent for publication

Not applicable. All data in this study are sourced from publicly available datasets.

Competing interests

We have no conflict of interest to declare. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Funding

This work was supported by the National Natural Science Foundation of China (No. 82302872); the Changchun Science and Technology Planning Project (No. 21ZY18).

Authors' contributions

Zhao TY: Methodology, formal analysis, data curation, visualization, and writing—original draft preparation; Li H: software; Zhang MS, Xu Y and Zhang M: writing—editing; Chen L: conceptualization and supervision. All authors have read and approved the published version of the manuscript.

Acknowledgements

We would like to thank Taylor & Francis (*www.tandfeditingservices.com*) for English language editing.

Availability of data and material

Data availability

Publicly available datasets were analyzed in this study. These data can be found here: [MRC IEU OpenGWAS] at (*https://gwas.mrcieu.ac.uk/*), and [UniProt] at (*https://www.uniprot.org/*), the above database search was completed on January 30, 2023.

Code availability

The MRAD platform can be freely accessed online at *https://gwasmrad.com/mrad/*.

The main project development repository: *https://github.com/ZhaoTianyu-zty/MRAD*.

Abbreviations

- AD: Alzheimer's disease
- APP: amyloid precursor protein
- Aβ: amyloid-β
- CI: confidence interval
- EOAD: early-onset Alzheimer's disease
- eQTL: expression Quantitative Trait Loci
- G protein: guanine nucleotide-binding protein G(k) subunit alpha
- GDP: guanosine diphosphate
- Gln: glutamine
- GTP: guanosine triphosphate
- GWAS: genome-wide association study
- HLA: human leukocyte antigen
- IVW: Inverse Variance Weighted
- LD: linkage disequilibrium
- LDL: low density lipoprotein
- LOAD: late-onset Alzheimer's disease
- MR: Mendelian randomization
- MRAD: Mendelian randomization for Alzheimer's disease
- OR: odds ratio
- PSME1: proteasome activator complex subunit 1
- RCT: randomized controlled trial
- SIGLECS: sialic acid-binding immunoglobulin like lectins
- SNPs: Single nucleotide polymorphisms
- SNXs: Sorting Nexins
- TBCA: tubulin-specific chaperone A
- TBCs: tubulin-specific chaperones
- VLDL: very low density lipoprotein
- VPS29: vacuolar protein sorting-associated protein 29
- WHO: World Health Organization

Highlights

(1) To the best of our knowledge, this is one of the most comprehensive studies to provide important insight into genetic etiology underlying AD based on hypothesis-free Mendelian randomization analysis. We generated 400,274 data entries in total, among which the major analysis method of IVW model consists of 73,129 records with 4840 exposure traits, which fall into 10 categories: Disease (n=17,168), Medical laboratory science (n=15,416), Imaging (n=4,896), Anthropometric (n=4,478), Treatment (n=4,546), Molecular trait (n=17,757), Gut microbiota (n=48), Past history (n=668), Family history (n=1,114), and Lifestyle trait (n=7,038).

(2) It is also important to note that we developed the first MR platform for AD, of great clinical and scientific significance that provided a thorough and comprehensive evaluation with regard to risk or protective factors for AD. It also provided physicians and scientists with a very convenient, free as well as user-friendly tool for further scientific investigation. The overall method used to construct this platform can be applied to the research of other diseases' etiology.

(3) It is also worth noting that we identified CD33, TBCA, VPS29, GNAI3, and PSME1 as novel potential therapeutic targets, which might be promising drug targets for AD and warrant further clinical investigation, especially TBCA and VPS29.

References

[1] J, G., J, J., D, P., et al. (2016) **Therapeutic strategies for Alzheimer's disease in clinical trials** *Pharmacol Rep* **68**

[2] Ransohoff R.M, EL Khoury J. (2015) **Microglia in Health and Disease** *Cold Spring Harb. Perspect. Biol* **8**

[3] Zou Z., Liu C., Che C., et al. (2014) **Clinical genetics of Alzheimer's disease** *BioMed Res. Int.* **2014**

[4] GBD 2019 DEMENTIA FORECASTING COLLABORATORS (2022) **Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the Global Burden of Disease Study 2019** *Lancet Public Health* **7**:e105–e125

[5] (2022) **2022 Alzheimer's disease facts and figures** *Alzheimers Dement* **18**:700–789

[6] Gauthier S, Webster C, Servaes S, Morais JA, Rosa-Neto P (2022) **World Alzheimer Report 2022 – Life after diagnosis: Navigating treatment, care and support**

[7] Falsetti L., Viticchi G., Buratti L., et al. (2018) **Interactions between Atrial Fibrillation, Cardiovascular Risk Factors, and ApoE Genotype in Promoting Cognitive Decline in Patients with Alzheimer's Disease: A Prospective Cohort Study** *J. Alzheimers Dis* **62**:713–725

[8] Femminella G.D., Taylor-Davies G., Scott J., et al. (2018) **Do Cardiometabolic Risk Factors Influence Amyloid, Tau, and Neuronal Function in APOE4 Carriers and Non-Carriers in Alzheimer's Disease Trajectory?** *J. Alzheimers Dis.* **64**:981–993

[9] Pegueroles J., Jiménez A., Vilaplana E., et al. (2018) **Obesity and Alzheimer's disease, does the obesity paradox really exist? A magnetic resonance imaging study** *Oncotarget* **9**:34691–34698

[10] Anstey K.J., Cherbuin N., Herath P.M (2013) **Development of a new method for assessing global risk of Alzheimer's disease for use in population health approaches to prevention** *Prev. Sci* **14**:411–421

[11] Jayaraman A., Pike C.J (2014) **Alzheimer's disease and type 2 diabetes: multiple mechanisms contribute to interactions** *Curr. Diab. Rep* **14**

[12] Vagelatos N.T., Eslick G.D (2013) **Type 2 diabetes as a risk factor for Alzheimer's disease: the confounders, interactions, and neuropathology associated with this relationship[J]** *Epidemiol. Rev* **35**:152–160

[13] Cooper J., Pastorello Y., Slevin M (2023) **A meta-analysis investigating the relationship between inflammation in autoimmune disease, elevated Crp, and the risk of dementia** *Front. Immunol* **14**

[14] Xu Q., Zhang Y., Zhang X., et al. (2020) **Medium-chain triglycerides improved cognition and lipid metabolomics in mild to moderate Alzheimer's disease patients with APOE4-/-: A double-blind, randomized, placebo-controlled crossover trial** *Clin. Nutr* **39**:2092–2105

[15] Zhu Y., Liu X., Zhu R., et al. (2022) **Lipid levels and the risk of dementia: A dose-response meta-analysis of prospective cohort studies** *Ann. Clin. Trans. Neurol* **9**:296–311

[16] Lopes Da Silva S., Vellas B., Elemans S., et al. (2014) **Plasma nutrient status of patients with Alzheimer's disease: Systematic review and meta-analysis** *Alzheimers Dement* **10**:485–502

[17] Yu J.T., Xu W., Tan C.C., et al. (2020) **Evidence-based prevention of Alzheimer's disease: systematic review and meta-analysis of 243 observational prospective studies and 153 randomised controlled trials** *J. Neurol. Neurosurg. Psychiatry* **91**:1201–1209

[18] Douaud G., Refsum H., De Jager C.A., et al. (2013) **Preventing Alzheimer's disease-related gray matter atrophy by B-vitamin treatment** *Proc. Natl. Acad. Sci. U.S.A* **110**:9523–9528

[19] Kivipelto M., Mangialasche F., Ngandu T (2018) **Lifestyle interventions to prevent cognitive impairment, dementia and Alzheimer disease** *Nat. Rev. Neurol* **14**:653–666

[20] Xu W., Tan L., Wang H.F., et al. (2015) **Meta-analysis of modifiable risk factors for Alzheimer's disease** *J. Neurol. Neurosurg. Psychiatry* **86**:1299–1306

[21] Livingston G., Sommerlad A., Orgeta V., et al. (2017) **Dementia prevention, intervention, and care** *Lancet* **390**:2673–2734

[22] Livingston G., Huntley J., Sommerlad A., et al. (2020) **Dementia prevention, intervention, and care: 2020 report of the Lancet Commission** *Lancet* **396**:413–446

[23] Hu Y., Zhang Y., Zhang H., et al. (2022) **Mendelian randomization highlights causal association between genetically increased C-reactive protein levels and reduced Alzheimer's disease risk** *Alzheimers Dement* **18**:2003–2006

[24] Williams D.M., Finan C., Schmidt A.F., et al. (2020) **Lipid lowering and Alzheimer disease risk: A mendelian randomization study** *Ann. Neurol* **87**:30–39

[25] Cui G., Li S., Ye H., Yang Y., Huang Q., Chu Y., Shi Z., Zhang X (2022) **Are neurodegenerative diseases associated with an increased risk of inflammatory bowel disease? A two-sample Mendelian randomization study** *Frontiers in immunology* **13**

[26] Hingorani A., Humphries S (2005) **Nature's randomised trials** *Lancet* **366**:1906–1908

[27] Emdin C. A., Khera A. V., Kathiresan S (2017) **Mendelian Randomization** *Jama* **318**:1925–1926

[28] Davey Smith G., Hemani G. (2014) **Mendelian randomization: genetic anchors for causal inference in epidemiological studies** *Hum. Mol. Genet* **23**:R89–98

[29] Davies N. M., Holmes M. V., Davey Smith G (2018) **Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians** *BMJ (Clinical research ed* **362**

[30] Chen Sd, Zhang W, Li Yz, et al. (2023) **A Phenome-wide Association and Mendelian Randomization Study for Alzheimer's Disease: A Prospective Cohort Study of 502, 493 Participants From the UK Biobank** *Biol Psychiatry* **93**:790–801

[31] Elsworth B., Lyon M., Alexander T., et al. (2020) **The MRC IEU OpenGWAS data infrastructure** *bioRxiv*

[32] Uniprot Consortium T. (2018) **UniProt: the universal protein knowledgebase** *Nucleic Acids Res* **46**

[33] Chen T., Zhang H., Liu Y., et al. (2021) **EVenn: Easy to create repeatable and editable Venn diagrams and Venn networks online** *J. Genet. Genomics* **48**:863–866

[34] (2022) **R: The R Project for Statistical Computing [EB/OL]. [2022-02-04]. https://www.r-project.org/.**

[35] Bowden J., Del Greco M.F., Minelli C., et al. (2017) **A framework for the investigation of pleiotropy in two-sample summary data Mendelian randomization** *Stat. Med* **36**:1783–1802

[36] Burgess S., Thompson S.G (2017) **Interpreting findings from Mendelian randomization using the MR-Egger method** *Eur. J. Epidemiol* **32**:377–389

[37] Slob E.A.W., Groenen P.J.F., Thurik A.R., et al. (2017) **A note on the use of Egger regression in Mendelian randomization studies** *Int. J. Epidemiol* **46**:2094–2097

[38] Bowden J., Davey Smith G., Haycock P.C., et al. (2016) **Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator** *Genet. Epidemiol* **40**:304–314

[39] Xue H., Shen X., Pan W (2021) **Constrained maximum likelihood-based Mendelian randomization robust to both correlated and uncorrelated pleiotropic effects** *Am. J. Human Genet* **108**:1251–1269

[40] JP H, SG T, JJ D, et al. (2003) **Measuring inconsistency in meta-analyses** *Bmj* **327**

[41] Bowden J., Davey Smith G., Burgess S (2015) **Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression** *Int. J. Epidemiol* **44**:512–525

[42] Hemani G., Zheng J., Elsworth B., et al. (2018) **The MR-Base platform supports systematic causal inference across the human phenome** *eLife* **7**

[43] (2023) **Shiny [EB/OL]. [2023-03-11]. https://shiny.rstudio.com/.**

[44] Picard C., Nilsson N., Labonté A., et al. (2022) **Apolipoprotein B is a novel marker for early tau pathology in Alzheimer's disease** *Alzheimers Dement* **185**:875–887

[45] Caramelli P., Nitrini R., Maranhão R., et al. (1999) **Increased apolipoprotein B serum concentration in Alzheimer's disease** *Acta Neurol. Scand* **100**:61–63

[46] et al. (1998) **Elevated low-density lipoprotein in Alzheimer's disease correlates with brain abeta 1-42 levels** *Biochem. Biophys. Res. Commun* **252**

[47] et al. (2006) **Association between apolipoprotein E polymorphism and serum lipid and apolipoprotein levels with Alzheimer's disease** *Neurosci. Lett* **408**

[48] et al. (2009) **Three-dimensional colocalization analysis of plasma-derived apolipoprotein B with amyloid plaques in APP/PS1 transgenic mice** *Histochem. Cell. Biol* **131**

[49] Takechi R., Galloway S., Pallebage-Gamarallage M.M.S., et al. (2010) **Differential effects of dietary fatty acids on the cerebral distribution of plasma-derived apo B lipoproteins with amyloid-beta. Brit** *J. Nutr* **103**:652–662

[50] Liu Y., Zhong X., Shen J., et al. (2020) **Elevated serum TC and LDL-C levels in Alzheimer's disease and mild cognitive impairment: A meta-analysis study** *Brain Res* **1727**

[51] Wu Y., Wang Z., Jia X., et al. (2019) **Prediction of Alzheimer's disease with serum lipid levels in Asian individuals: a meta-analysis** *Biomarkers* **24**:341–351

[52] Wang P., Zhang H., Wang Y., et al. (2020) **Plasma cholesterol in Alzheimer's disease and frontotemporal dementia** *Transl. Neurosci* **11**:116–123

[53] Zhou Z., Liang Y., Zhang X., et al. (2020) **Low-Density Lipoprotein Cholesterol and Alzheimer's Disease: A Systematic Review and Meta-Analysis** *Front. Aging Neurosci* **12**

[54] Nordestgaard L.T., Christoffersen M., Frikke-Schmidt R (2022) **Shared Risk Factors between Dementia and Atherosclerotic Cardiovascular Disease** *Int. J. Mol. Sci* **23**

[55] Wingo A.P., Vattathil S.M., Liu J., et al. (2022) **LDL cholesterol is associated with higher AD neuropathology burden independent of APOE** *J. Neurol. Neurosurg. Psychiatry* **93**:930–938

[56] Hartmann T (2001) **Cholesterol, A beta and Alzheimer's disease** *Trends Neurosci* **24**:S45–48

[57] Tynkkynen J., Chouraki V., Van Der Lee S.J., et al. (2018) **Association of branched-chain amino acids and other circulating metabolites with risk of incident dementia and Alzheimer's disease: A prospective study in eight cohorts** *Alzheimer Dement* **14**:723–733

[58] Hashemi M., Banerjee S., Lyubchenko Y.L (2022) **Free Cholesterol Accelerates Aβ Self-Assembly on Membranes at Physiological Concentration** *Int. J. Mol. Sci* **23**

[59] Treiber-Held S., Distl R., Meske V., et al. (2003) **Spatial and temporal distribution of intracellular free cholesterol in brains of a Niemann-Pick type C mouse model showing hyperphosphorylated tau protein. Implications for Alzheimer's disease** *J. Pathol* **200**:95–103

[60] Zarrouk A., Debbabi M., Bezine M., et al. (2018) **Lipid Biomarkers in Alzheimer's Disease** *Curr. Alzheimer Res* **15**:303–312

[61] Zuin M., Cervellati C., Trentini A., et al. (2021) **Association between Serum Concentrations of Apolipoprotein A-I (ApoA-I) and Alzheimer's Disease: Systematic Review and Meta-Analysis** *Diagnostics* **11**

[62] Tong J.H., Gong S.Q., Zhang Y.S., et al. (2022) **Association of Circulating Apolipoprotein AI Levels in Patients with Alzheimer's Disease: A Systematic Review and Meta-Analysis[J]** *Front. Aging Neurosci* **14**

[63] Button E.B., Boyce G.K., Wilkinson A., et al. (2019) **ApoA-I deficiency increases cortical amyloid deposition, cerebral amyloid angiopathy, cortical and hippocampal astrogliosis, and amyloid-associated astrocyte reactivity in APP/PS1 mice** *Alzheimers Res. Ther* **11**

[64] Cannon-Albright L.A., Foster N.L., Schliep K., et al. (2019) **Relative risk for Alzheimer disease based on complete family history** *Neurology* **92**:e1745–e1753

[65] Wang Y., Wang Q., Li J., et al. (2019) **Glutamine Improves Oxidative Stress through the Wnt3a/β-Catenin Signaling Pathway in Alzheimer's Disease In Vitro and In Vivo** *BioMed Res. Int* **2019**

[66] Letenneur L., Commenges D., Dartigues J.F., et al. (1994) **Incidence of dementia and Alzheimer's disease in elderly community residents of south-western France** *Int. J. Epidemiol* **23**:1256–1261

[67] Y S, B G, TK T, et al. (1994) **Influence of education and occupation on the incidence of Alzheimer's disease** *Jama* **271**

[68] Illegas-Llerena C., Phillips A., Garcia-Reitboeck P., et al. (2016) **Microglial genes regulating neuroinflammation in the progression of Alzheimer's disease** *Curr. Opin. Neurobiol* **36**:74–81

[69] Malik M., Simpson J.F., Parikh I., et al. (2013) **CD33 Alzheimer's risk-altering polymorphism, CD33 expression, and exon 2 splicing** *J. Neurosci* **33**:13320–13325

[70] Hollingworth P., Harold D., Sims R., et al. (2011) **Common variants at Abca7, MS4A6A/Ms4a4e, Epha1, CD33 and CD2AP are associated with Alzheimer's disease** *Nat. Genet* **43**:429–435

[71] Gu X, Dou M, Cao B, Jiang Z, Chen Y (2022) **Peripheral level of CD33 and Alzheimer's disease: a bidirectional two-sample Mendelian randomization study** *Transl Psychiatry* **12**

[72] et al. (2013) **Alzheimer's disease risk gene CD33 inhibits microglial uptake of amyloid beta** *Neuron* **78**

[73] Wang W.Y., Liu Y., Wang H.F., et al. (2017) **Impacts of CD33 Genetic Variations on the Atrophy Rates of Hippocampus and Parahippocampal Gyrus in Normal Aging and Mild Cognitive Impairment** *Mol. Neurobiol* **54**:1111–1118

[74] Miles L.A., Hermans S.J., Crespi G.A.N., et al. (2019) **Small Molecule Binding to Alzheimer Risk Factor CD33 Promotes Aβ Phagocytosis** *iScience* **19**:110–118

[75] Nolasco S., Bellido J., Gonçalves J., et al. (2005) **Tubulin cofactor A gene silencing in mammalian cells induces changes in microtubule cytoskeleton, cell cycle arrest and cell death** *FEBS Lett* **579**:3515–3524

[76] Gaitanos T.N., Buey R.M., Díaz J.F., et al. (2004) **Peloruside A does not bind to the taxoid site on beta-tubulin and retains its activity in multidrug-resistant cell lines** *Cancer Res* **64**:5063–5067

[77] Cormier A., Marchand M., Ravelli R.B.G., et al. (2008) **Structural insight into the inhibition of tubulin by vinca domain peptide ligands** *EMBO Rep* **9**:1101–1106

[78] Bellmunt J., Théodore C., Demkov T., et al. (2009) **Phase III trial of vinflunine plus best supportive care compared with best supportive care alone after a platinum-containing regimen in patients with advanced transitional cell carcinoma of the urothelial tract** *J. Clin. Oncol* **27**:4454–4461

[79] Hillary R.F., Gadd D.A., Mccartney D.L., et al. (2022) **Genome- and epigenome-wide studies of plasma protein biomarkers for Alzheimer's disease implicate TBCA and TREM2 in disease risk** *Alzheimers Dement* **14**

[80] Ye H., Ojelade S.A., Li-Kroeger D., et al. (2020) **Retromer subunit, Vps29, regulates synaptic transmission and is required for endolysosomal function in the aging brain** *eLife* **9**

[81] Fuse A., Furuya N., Kakuta S., et al. (2015) **VPS29-VPS35 intermediate of retromer is stable and may be involved in the retromer complex assembly process** *FEBS Lett* **589**:1430–1436

[82] Jimenez-Orgaz A., Kvainickas A., Nägele H., et al. (2018) **Control of RAB7 activity and localization through the retromer-TBC1D5 complex enables RAB7-dependent mitophagy** *EMBO J* **37**:235–254

[83] Zhang Q.Y., Tan M.S., Yu J.T., et al. (2016) **The Role of Retromer in Alzheimer's Disease** *Mol. Neurobiol* **53**:4201–4209

[84] Seaman M.N.J (2021) **The Retromer Complex: From Genesis to Revelations** *Trends Biochem. Sci* **46**:608–620

[85] Heese K (2013) **G proteins, p60trp, and neurodegenerative diseases** *Mol. Neurobiol.* **47**:1103–1111

[86] Fowler C.J., Cowburn R.F., Garlind A., et al. (1995) **Disturbances in signal transduction mechanisms in Alzheimer's disease** *Mol. Cell. Biochem* **149**:287–292

[87] Adelöf J, Andersson M, Porritt M, et al. (2018) **PA28αβ overexpression enhances learning and memory of female mice without inducing 20S proteasome activity** *BMC neuroscience* **19**

[88] Donggui WU. (2021) **Study on the role of PA28 in protein metabolic pathway of oxidative damage in the brain of diabetic rats** *Dali University* :1–52

Editors

Reviewer #1 (Public Review):

Summary:

An online database called MRAD has been developed to identify the risk or protective factors for AD.

Strengths:

This study is a very intriguing study of great clinical and scientific significance that provided a thorough and comprehensive evaluation with regard to risk or protective factors for AD. It also provided physicians and scientists with a very convenient, free as well as user-friendly tool for further scientific investigation.

Comments on revised version:

The authors have resolved all of my previous comments. It's a decent paper worth to be published in this field.

https://doi.org/10.7554/eLife.96224.2.sa2

Reviewer #2 (Public Review):

Summary:

This MR study by Zhao et al. provides a comprehensive hypothesis-free approach to identifying risk and protective factors causal to Alzheimer's Disease (AD).

Strengths:

The study employs a comprehensive, hypothesis-free approach, which is novel over traditional hypothesis-driven studies. Also, causal associations between risk/protective factors and AD were addressed using genetic instruments and analysis.

https://doi.org/10.7554/eLife.96224.2.sa1

Author response:

The following is the authors' response to the original reviews.

> ***Public Reviews:***
>
> ***Reviewer #1 (Public Review):***
>
> *Summary:*
>
> *An online database called MRAD has been developed to identify the risk or protective factors for AD.*
>
> *Strengths:*
>
> *This study is a very intriguing study of great clinical and scientific significance that provided a thorough and comprehensive evaluation with regard to risk or protective factors for AD. It also provided physicians and scientists with a very convenient, free as well as user-friendly tool for further scientific investigation.*

We thank the reviewer for the conclusion and positive comments.

> *Weaknesses:*
>
> *(1) Comment: The paper mentions that the MRAD database currently contains data only from European populations, with no mention of data from other populations or ethnicities. Given potential differences in Alzheimer's Disease (AD) across different populations, the limitations of the data should be emphasized in the discussion, along with plans to expand the database to include data from more racial and geographic regions.*

Thank you for your valuable comment. Further information regarding the limitations of populations is provided in the Conclusions section (page 19).

The newly added text describing the limitations of populations is as follows:

"However, in this study, since the GWAS datasets for both the exposure and the outcome traits (AD) selected were obtained from the public database (MRC IEU OpenGWAS), where the GWAS datasets for AD are only of European population, and since we use the TwoSampleMR, which requires that the populations for the exposure trait and the outcome trait be the same to satisfy the requirement for a control variable, this study currently has certain limitations in terms of population. We initiated a Mendelian randomization study on AD at clinical hospitals in China and are currently in the sample collection stage to address the limitations. In the future, we will integrate data from more populations and keep updating new progresses in AD research to explore its potential differences in different populations."

> *(2) Comment: Sufficient information should be provided to clarify the data sources, sample selection, and quality control methods used in the MRAD database. Readers may expect more detailed information about the data to ensure data reliability, representativeness, and research applicability.*

Thank you for your helpful suggestion. We appreciate you taking time and making effort in reviewing our manuscript and thank you for your insightful comments. We agree that adding more details is essential to make the manuscript more reliability, representativeness, and research applicability.

The newly added text describing more detailed information about the data is as follows:

(1) Sufficient information about data sources and sample selection (in the Data sources section of Methods section, page 8):

"Exposure traits

Inclusion criteria: datasets of the European population.

Exclusion criteria: (i) eQTL-related datasets; (ii) AD-related datasets.

"In this study, the GWAS datasets selected were derived from 42,335 GWAS datasets in the public database (MRC IEU OpenGWAS, https://gwas.mrcieu.ac.uk/). Based on the above inclusion and exclusion criteria, 19,942 eQTL-related datasets were excluded first, leaving 22,393 GWAS datasets. Next, the datasets with the European population were selected, and 18,117 GWAS datasets were obtained. Finally, 20 AD-related datasets were excluded; 18,097 GWAS datasets were obtained at the end as the exposure traits of this study (See Table S1 for basic information).

Outcome traits

Inclusion criteria: (i) datasets of patients with AD with complete information and clear data sources; (ii) datasets of the European population.

Exclusion criteria: (i) Number of SNPs <1 million; (ii) datasets with unspecified sex; (iii) datasets with a family history of AD; (iv) datasets with dementia.

Based on the above criteria, 16 GWAS datasets of outcome traits were selected from the MRC IEU OpenGWAS database, comprising datasets of AD from Alzheimer Disease Genetics Consortium (ADGC), Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium (CHARGE), The European Alzheimer's Disease Initiative (EADI), and Genetic and Environmental Risk in AD/Defining Genetic, Polygenic and Environmental Risk for Alzheimer's Disease Consortium (GERAD/PERADES) 2019 (ieu-b-2); AD from Benjamin Woolf 2022 (ieu-b-5067); AD from International Genomics of Alzheimer's Project (IGAP) 2013 (ieu-a-297) as the datasets of main outcome traits for AD, as well as 13 datasets from FinnGen

biobank 2021 corresponding to various AD subtypes, referred to as AD-finn subtypes. (as shown in Figure 2)."

(2) Sufficient information about quality control methods (in the Statistical models for causal effect inference section of Methods section, page 9-10:

"A random-effects IVW model was used in this study as the major analysis method to uncover potential risk or protective factors for AD. The random-effects IVW model as the gold standard for MR studies, its principle is to calculate the inverse of the variance of each IV as its weight, assuming all IVs are valid. The regression does not include an intercept term, and the final result is the weighted average of the effect estimates from all IVs [34]. This model indicates that the true effect values may vary across different studies due to both sampling error and the heterogeneity of the true effect. The weight of each study is jointly determined by its inverse variance and the estimated heterogeneity variance. Thus, as long as there is no pleiotropy, even when there is significant heterogeneity ($p < 0.05$), this method remains the best MR model.

To assess the robustness of the IVW results, sensitivity analysis was performed using six additional models: (i) MR-Egger: MR-Egger's biggest difference from IVW is that it considers the intercept term during regression to evaluate bias caused by horizontal pleiotropy. The intercept represents the magnitude of horizontal pleiotropy, with a value close to 0 indicating minimal pleiotropy. The primary purpose is to detect and correct for horizontal pleiotropy. Thus, when significant horizontal pleiotropy is observed ($p < 0.05$), this method is preferred [35,36]. (ii) Weighted median: The weighted median method is a technique for evaluating causal relationships using a majority of genetic variants (SNPs). If at least 50% of the SNPs are valid IVs, the median of the causal estimates will tend toward the true causal effect. This method provides an unbiased estimate (i.e., the "majority validity" assumption) [37]. (iii) Simple mode: Involves comparing the frequencies or proportions of genotypes or phenotypes between control and experimental groups. Moreover, it can illustrate whether the observed differences in genotypes or phenotypes between the two groups are statistically significant. (iv) Weighted mode: The weighted mode method is a technique for combining multiple Mendelian randomization estimates. This method assigns weights to the causal effect estimates of different genetic variants on the trait and then takes the weighted mode as the final estimate of the causal effect. In genetic variant estimates, the method can decrease bias caused by outliers. (v) Maximum likelihood: This method is used when it is known that a random sample follows a particular probability distribution; however, the specific parameters of that distribution remain unknown, and it involves conducting multiple experiments, observing the results, and using those results to infer the approximate values of the parameters [38]. (vi) Penalized weighted median: An enhanced version of the weighted median estimate that provides a consistent estimate of the causal effect. (vii) Heterogeneity and horizontal pleiotropy assessment use the heterogeneity tests [39] and Egger intercept tests [40], respectively."

> *(3) Comment: While the authors mention that the MRAD database offers interactive visualization interfaces, the paper lacks detailed information on how to interpret and understand these visual results. Guidelines on effectively using these visualization tools to help researchers better comprehend the data are essential.*

Thank you very much for your feedback, as we believe that our manuscript has been improved substantially as a result of your input. Owing to space constraints, the MRAD database user guide is included in the Supplementary Material. Meanwhile, for better understanding, the subheading of the relevant content in the Supplementary Material has been revised to "MRAD User Guide" (see Supplementary Material for details, page 11). Furthermore, considering user-friendliness, the user guide has been integrated into the

database and can be accessed directly from the homepage by clicking on the "User Guide" module.

> *(4) Comment: In the conclusion section of the paper, it is advisable to explicitly emphasize the practical applications and potential clinical significance of the MRAD database. The paper should articulate how MRAD can contribute to the early identification, diagnosis, prevention, and treatment of AD and its potential societal and clinical value more clearly.*

Thank you for pointing this out. In the Discussion section of the revised manuscript, we have now added how MRAD can contribute to the early identification, diagnosis, prevention, and treatment of AD and its potential societal as well as clinical value. And we reorganized the structure of Discussion section to make the text easier to understand, which could be helpful to further clarify the significance of MRAD. (page 15)

The newly added text describing the practical applications and potential clinical significance of the MRAD database is as follows:

"(i) The current methods for identifying AD mainly rely on assessment scales, cerebrospinal fluid (CSF) examinations, and brain PET/MRI. However, assessment scales can be biased by factors such as the anxiety and nervousness of the subjects. CSF examinations require an invasive lumbar puncture, leading to low patient acceptance. PET/MRI scans are expensive and have limited equipment accessibility. These limitations restrict early AD identification. Thus, there is a pressing clinical need for readily available, time- and cost-effective, and accurate detection methods. In this study, the Medical laboratory science and Molecular trait used could be less expensive, faster to detect, easier to operate, and more accessible for widespread adoption. They hold great value for early AD identification and have the potential to become crucial tools for identifying AD in the future. (ii) Imaging acts as a powerful assistive tool for diagnosing Alzheimer's disease. Traditional imaging examinations mainly depict changes in the brain's macroscopic structure, while research on microstructural changes in disease-related areas is relatively limited. Studies have demonstrated that microstructural neurodegenerative processes are extensive and pronounced during AD progression. Our study results cover traditional macroscopic neuroimaging results and reveal numerous potential causal relationships between brain microstructure and AD. The combination of macroscopic and microstructural insights will provide more valuable information for clinical diagnosis. (iii) Clarifying patient's disease, past history, and family history can aid in preventing AD at an early stage, and prevention of AD could be attained through monitoring anthropometric indicators, improving gut microbiota, and adjusting lifestyle traits. (iv) Currently, the development of new drugs for AD is mainly underscored by Aβ, Tau, and other inhibitors. Since 2000, global pharmaceutical companies have invested hundreds of billions of dollars in the development of new drugs for AD, and these drugs have not yielded successful results. AD drug development has thus been perceived as having the highest failure rate of all drug research, reaching 99.6%. Hence, further research on molecular traits to find new targets and develop new drugs for these targets will provide new pathways for AD treatment."

> *(5) Comment: Grammar and Spelling Errors: There are several spelling and grammar errors in the paper. Referring to a scientific editing service is recommended.*

We appreciate your comments and suggestions for improving our manuscript. We have now used a professional editing service offered by Taylor and Francis to revise the grammar and language, and we have obtained a certificate of proof, which is attached. Thank you for recognizing our research, we have tried our best to improve the quality of this paper to ensure that it meets the high standards required for publication in of journal elife.

> ***Reviewer #2 (Public Review):***
>
> *Summary:*
>
> *This MR study by Zhao et al. provides a comprehensive hypothesis-free approach to identifying risk and protective factors causal to Alzheimer's Disease (AD).*
>
> *Strengths:*
>
> *The study employs a comprehensive, hypothesis-free approach, which is novel over traditional hypothesis-driven studies. Also, causal associations between risk/protective factors and AD were addressed using genetic instruments and analysis.*

We greatly appreciate the positive feedback regarding the overall quality of our work.

> *Major comments:*
>
> *(1) Comment: The authors used the inverse-variance weighted (IVW) model as the primary method and other MR methods (MR-Egger, weighted mean, etc.) for sensitivity analysis. However, each method has its own assumption, and IVW is only robust when pleiotropy and heterogeneity are not severe. Rather than using IVW imprudently across all associations, it would be more appropriate to choose the best MR method for each association based on heterogeneity/Egger intercept tests. This customized approach, based on tests of MR assumption violations, yields more stable and reliable results. For reference, please follow up on work by Milad et al. (EHJ - "Plasma lipids and risk of aortic valve stenosis: a Mendelian randomization study"). This study selected the best MR model for each association based on pleiotropy and heterogeneity tests. Given the large number of tests in this work, I suggest initially screening significant signals using IVW, as done, and then validating the results using multiple MR methods for those signals. It is common for MR estimates from different methods to vary significantly (with some being statistically significant and others not), and in such cases, the MR estimates from the best-fitted model should be trusted and highlighted.*

Thank you for your professional comments. We agree that our description of the Statistical models for causal effect inference was not specific enough. Therefore, we have included a new text describing more details about each method's assumption and supplied a predefined approach to select the best statistical estimation from these methods in the Statistical models for causal effect inference section of Methods section (page 9-10). However, we would like to clarify our analysis method. In this study, the main analysis method used is the IVW random effects model instead of the IVW fixed effects model. The IVW random effects model indicates that the true effect values of different studies may vary, including both sampling error and heterogeneity of the true effect. The weight of each study is jointly determined by its inverse variance and the estimated heterogeneity variance. Thus, as long as there is no pleiotropy, even when there is significant heterogeneity ($p < 0.05$), this method is still the best MR model. We would like to thank you again for your feedback, as we believe that our manuscript has been improved substantially as a result of your input.

The newly added text describing more details about each method's assumption and the customized best-fitted model is as follows:

"Statistical models for causal effect inference

A random-effects IVW model was used in this study as the major analysis method to uncover potential risk or protective factors for AD. The random-effects IVW model as the gold standard for MR studies, its principle is to calculate the inverse of the variance of each IV as its weight, assuming all IVs are valid. The regression does not include an intercept term, and the final result is the weighted average of the effect estimates from all IVs [34]. This model

indicates that the true effect values may vary across different studies due to both sampling error and the heterogeneity of the true effect. The weight of each study is jointly determined by its inverse variance and the estimated heterogeneity variance. Thus, as long as there is no pleiotropy, even when there is significant heterogeneity ($p < 0.05$), this method remains the best MR model.

To assess the robustness of the IVW results, sensitivity analysis was performed using six additional models: (i) MR-Egger: MR-Egger's biggest difference from IVW is that it considers the intercept term during regression to evaluate bias caused by horizontal pleiotropy. The intercept represents the magnitude of horizontal pleiotropy, with a value close to 0 indicating minimal pleiotropy. The primary purpose is to detect and correct for horizontal pleiotropy. Thus, when significant horizontal pleiotropy is observed ($p < 0.05$), this method is preferred [35,36]. (ii) Weighted median: The weighted median method is a technique for evaluating causal relationships using a majority of genetic variants (SNPs). If at least 50% of the SNPs are valid IVs, the median of the causal estimates will tend toward the true causal effect. This method provides an unbiased estimate (i.e., the "majority validity" assumption) [37]. (iii) Simple mode: Involves comparing the frequencies or proportions of genotypes or phenotypes between control and experimental groups. Moreover, it can illustrate whether the observed differences in genotypes or phenotypes between the two groups are statistically significant. (iv) Weighted mode: The weighted mode method is a technique for combining multiple Mendelian randomization estimates. This method assigns weights to the causal effect estimates of different genetic variants on the trait and then takes the weighted mode as the final estimate of the causal effect. In genetic variant estimates, the method can decrease bias caused by outliers. (v) Maximum likelihood: This method is used when it is known that a random sample follows a particular probability distribution; however, the specific parameters of that distribution remain unknown, and it involves conducting multiple experiments, observing the results, and using those results to infer the approximate values of the parameters [38]. (vi) Penalized weighted median: An enhanced version of the weighted median estimate that provides a consistent estimate of the causal effect. (vii) Heterogeneity and horizontal pleiotropy assessment use the heterogeneity tests [39] and Egger intercept tests [40], respectively."

> *(2) Comment: Lines 157-160 mentioned "But to date, AD has been reported as hypothesis-driven MR study based on a single factor, ignoring the potential role of a huge number of other risk factors. Also, due to the high degree of heterogeneity present in AD subtypes, which have different biological and genetic characteristics. Thus, the previous studies cannot offer a systematic and complete viewpoint.". This statement overlooks a similar study published in Molecular Psychiatry ("A Phenome-wide Association and Mendelian Randomization Study for Alzheimer's Disease: A Prospective Cohort Study of 502,493"), which rigorously assessed the effects of 4171 factors spanning 10 different categories on AD using observational analysis and MR. The authors should revise their statement on the novelty of their study type throughout the manuscript and discuss how their work differs from and potentially strengthens previous studies.*

Thank you for directing us to this literature. We have read this article carefully. This study shares some similarities with our study but there are significant differences with regards to sample sources and research fields. The study, as mentioned by the reviewer, used the UKB database as its sample source, and analyzed the association between 10 categories (comprising 4,171 factors) and AD, which were sociodemographic, physical measures, lifestyle and environment, health conditions, mental health, medications and operations, cognitive function, sex-specific factors, employment, and early-life factors. However, the study revealed they are restricted by the available variables from the UKB database, which lead to variables such as air pollution, blood glucose measures and so on were not included. Conversely, our study used samples from the MRC IEU OpenGWAS database, the largest open

GWAS database globally. Furthermore, our research focus differs, as we primarily investigate the causal relationship between the following 10 categories (comprising 18,097 traits) and AD, which were Disease, Medical laboratory science, Imaging, Anthropometric, Treatment, Molecular trait, Gut microbiota, Past history, Family history, and Lifestyle trait. Most importantly, we have established a database encompassing all MR analysis results, allowing researchers and clinicians worldwide to conveniently and rapidly retrieve AD-associated risk factors via an online open integrated platform (MRAD, https://gwasmrad.com/mrad/).We have now added a new text in the Background section (page 6-7) describing the differences and potential strengthens towards previous studies.

The newly added text describing the differences and novelty towards previous studies is as follows:

“Chen et al. [30] used MR analysis to reveal the causal relationship between AD and factors including sociodemographic and early life status. However, the study revealed they are restricted by the available variables from the UKB database, which lead to variables such as air pollution, blood glucose measures and so on were not included. And also, due to the high degree of heterogeneity present in AD subtypes, which have different biological and genetic characteristics. Thus, the previous studies cannot offer a systematic and complete viewpoint. Our study uses the MRC IEU OpenGWAS database as the sample source for MR analysis to address the aforementioned limitations. The MRC IEU OpenGWAS database, the largest open GWAS database globally, has compiled 42,335 GWAS summary datasets from sources such as the UK Biobank, FinnGen Biobank, and Biobank Japan. Analyzing large-scale datasets will break new ground for MR research on AD.

MR requires a combination of background knowledge in biology, computer science, software studies, and statistics, which often leads to a dilemma where biologists are not well-versed in computer and statistical fields, while computer science experts struggle to adopt a medical biology mindset. Consequently, the vast majority of available GWAS data have not been effectively utilized through MR. Therefore, the construction of a multi-level data platform specifically for AD based on MR analysis of massive GWAS data is of great strategic significance, and it will facilitate researchers and clinicians worldwide to conveniently and rapidly obtain risk factors that are causally associated with AD.”

Reference:

[30] Chen SD, Zhang W, Li YZ, et al. (2023). A Phenome-wide Association and Mendelian Randomization Study for Alzheimer's Disease: A Prospective Cohort Study of 502,493 Participants From the UK Biobank. Biol Psychiatry. 1;93(9):790-801.

> *(3) Comment: Given the large number of tests, the multiple testing issue is concerning. To mitigate potential false positives, I recommend employing the Bonferroni threshold or FDR. The authors should only interpret exposures that are significant at the Bonferroni threshold.*

We sincerely appreciate the reviewer's feedback. Thank you for pointing this out. We have added the results of the Bonferroni correction to the Statistical models for the causal effect inference section of the Methods section (page 10) in response to the reviewer's feedback.

The newly added text describing Bonferroni threshold is as follows:

“The above analyses were performed using the TwoSampleMR[41] package in the R (version 4.1.2) software. Association of exposures with outcomes was assessed using odds ratio (OR) and 95% confidence interval (95% CI), with OR > 1 indicating a positive association (risk factor) and 0 < OR < 1 indicating a negative association (protective factor). Differences with a two-sided $p < .05$ were considered statistically significant. Furthermore, owing to the

relatively large number of exposure and outcome traits included in this study, the multiple testing correction method Bonferroni correction was added to identify significant hits, threshold for Bonferroni-corrected was 0.05 divided by 289,552 tests (p <1.727e-07)."

> *(4) Comment: In the discussion, the authors should interpret or highlight exposures that remain significant after multiple testing corrections.*

Thank you for your valuable comment. In response to reviewer feedback, we have put extra emphasis on the exposures that remained significant after multiple testing corrections in the Discussion section (page 17). We thank you again for your feedback, as we believe that our manuscript has been improved substantially as a result of your input.

> ***Recommendations for the authors:***
>
> ***Reviewer #1 (Recommendations For The Authors):***
>
> *(1) Comment: In this study, the authors used the inverse-variance weighted (IVW) model as the major analysis method to perform Mendelian randomization analysis to identify various classes of risk or protective factors for AD, early-onset AD, and late-onset AD. An online database called MRAD has been thereby developed with the assistance of Shiny package. This study is a very intriguing study of great clinical and scientific significance that provided a thorough and comprehensive evaluation with regard to risk or protective factors for AD. It also provided physicians and scientists with a very convenient, free as well as user-friendly tool for further scientific investigation.*
>
> *I believe this manuscript is great research that is worth publishing with all the comments from the Public Review resolved.*

We thank the reviewer for taking the time to read and provide valuable feedback on our manuscript, which allowed us to improve the overall quality of our research. All the comments from the Public Review have been rechecked, and appropriate changes have been made in accordance with the reviewers' suggestions. Point-by-point responses to all the comments from the Public Review can be found in the above. If there are any further issues, please do not hesitate to let us know, so that we can ensure that our manuscript meets the high standards required for publication.

> ***Reviewer #2 (Recommendations For The Authors):***
>
> *(1) Comment: In the middle lower left section of the graphical abstract, the overlapping positive (N=63) and overlapping negative (N=16) do not sum to the overlapping number (N=80). Could you clarify if any have both positive and negative effects? Additionally, the font size inside the circular elements is too small to read.*

We thank you for raising this issue. We have clarified this in the MRAD utility data mining section of Results section (page 12): A total of 63 exposure traits (risk factors) were positively associated with all the three main outcome traits, while 16 exposure traits (protective factors) were negatively associated with the three main outcome traits, with Ulcerative colitis (ebi-a-GCST000964) being negatively associated with the AD outcome traits of ieu-b-2 and ieu-a-297, and positively associated with the AD outcome traits of ieu-b-5067. Additionally, we apologize for the small, unreadable fonts in the graphical abstract figure. In response to reviewer feedback, we have increased the font size within the figure and enhanced the resolution to improve image readability (page 3).

> *(2) Comment: The x-axis label ("Alzheimer's disease outcome") should be more descriptive. If published GWAS results are used, indicate this as XXX et al. (2022). Also,*

> *specify the AD outcome for each category (e.g., AD, early-onset AD, late-onset AD). The y-axis labels should also be clarified; remove identification codes and retain only the exposure names. Apply the same improvements to Figures 2-8.*

We appreciate your comments and suggestions for improving our manuscript.

(i) In response to reviewer feedback, information of published GWAS such as authors and year of publication have now been added to the x-axis labels, as demonstrated in Figure 4 (page 31).

(ii) The outcome IDs are unique. We used these IDs to represent the AD information on the x-axis to maintain a clean and clear figure. The corresponding details for each ID are explained in the Outcome traits section of the Methods section (page 8, as shown in Figure 2). AD_EO refers to early-onset AD, and AD_LO refers to late-onset AD, which are also specified in the Abbreviations (page 4).

(iii) We sincerely appreciate the reviewers' meticulous feedback. While exposure IDs in this study are unique, exposure names are not. A single exposure name may correspond to multiple IDs, each with a potentially different source of information (e.g., author, year, population sample). We believe obtaining consistent results across multiple IDs further strengthens the reliability of our conclusions. Hence, for better clarity of specific exposure information, the exposure IDs have been retained.

> *(3) Comment: The results across Figures 1-8 are repetitive and not very informative. Consider other visualizations to condense the information into one or two figures. I would recommend using a Manhattan plot or PheWAS plot concept to effectively display many test results at once. Please display the Bonferroni threshold in the plot as a horizontal line to show which exposures are meaningful after adjusting multiple comparisons.*

We appreciate this helpful suggestion. We have now condensed Figures 1–8 into a single figure (as shown in Figure 4). Additionally, we have now displayed the Bonferroni correction results in the sensitivity analysis results figures (as shown in Figure 5, Figure S1-S7).

> *(4) Comment: Consider placing Figure S1 as Figure 1, condensing Figures 1-8 into Figures 2 and 3, and placing the circular diagrams from Figure S6 as Figure 4.*

We appreciate this valuable suggestion. The sequence of the figures has been adjusted.

> *(5) Comment: Create a main table summarizing robust and consistent exposures for AD that are significant at the Bonferroni threshold for readers. For each exposure, please include estimates from IVW, MR-Egger, weighted median, simple mode, weighted mode, maximum likelihood, and penalized weighted median, along with heterogeneity and horizontal pleiotropy tests. I would also highlight or bold estimates from the best-fit model/MR method to help readers identify the most reliable estimates when estimates from multiple methods are heterogeneous.*

We appreciate this helpful suggestion. Owing to the excessive amount of information in the table, we have uploaded the table covering the aforementioned information according to the reviewer's suggestion as supplementary materials (See Table S2). (i) The corresponding id.exposure that pass the Bonferroni threshold are reflected in red font. (ii) Furthermore, according to the customized best-fitted model (as mentioned in the Statistical models for causal effect inference section of Methods section), when there is no pleiotropy or when pleiotropy is not applicable (less than 3 SNPs), random-effects IVW model is the best model. These corresponding id.exposure are shown in red font with a yellow highlight. (iii)

Moreover, according to the customized best-fitted model, when there is pleiotropy, MR-Egger is the best model. These corresponding id.exposure are shown in red font with a green highlight.

> *(6) Comment: Figures S4-S10: These figures are screenshots of web browsers and may not be worth showing. Consider using tools like Adobe AI or R ggplot to create more refined visualizations that are specific to the research question and improve the message of this work.*

Thank you very much for your valuable suggestion in reviewing our manuscript. In this study, Figures S4-S10 are screenshots related to the user guide. We sincerely appreciate the reviewer's feedback and have revised the subheading of this section to MRAD User Guide to clarify its purpose. Demonstrating both text and figures in this section, we aim to help users understand ways to operate MRAD more intuitively and easily.

> *(7) Comment: Additionally, please show upfront or highlight results from MR analyses based on R packages, as the author mentioned in the method section. Somehow it's difficult to find results from MR-Egger, weighted median, simple mode, weighted mode, maximum likelihood, and penalized weighted median, along with heterogeneity and horizontal pleiotropy tests in the supplementary materials. Apologies if I missed them. Please ensure these results are clearly presented.*

We appreciate your comments and suggestions for improving our manuscript. Thank you for pointing this out. We have added the results of the sensitivity analysis based on R packages (as shown in Figure 5, Figure S1-S7, and Table S2).

https://doi.org/10.7554/eLife.96224.2.sa0