

LETTER TO THE EDITOR

Open Access



Identifying flaws in the GWAS datasets of a published Mendelian randomization study: complementary re-evaluation and suggestion for analytical refinements

Jia-Cheng Xiang^{1†}, Yi-Fan Xiong^{1†}, Shao-Gang Wang^{1*†}  and Qi-Dong Xia^{1*†}

To the editor,

In “Mendelian randomization and transcriptomic analysis reveal an inverse causal relationship between Alzheimer’s disease and cancer”, Zehua Dong and colleagues discovered a general protective effect of Alzheimer’s disease (AD) on cancer. However, after searching in a widely used GWAS database, IEU Open GWAS (<https://gwas.mrcieu.ac.uk/>) [1], we found several incorrect GWAS datasets were employed: ebi-a-GCST005921 [2] was used as exposure dataset for AD, which is actually “family history of AD”; furthermore, ukb-b-17001 and ukb-a-296 actually represent “ever had bowel cancer screening”, which were used as outcome datasets for bowel cancer (Fig. 1). Apparently, the authors used the incorrect GWAS datasets and did not explain for it. However, with a high heritability (60–80%), AD does have a strong correlation with AD family history [3]. We collated the recently published large AD GWAS dataset (ebi-a-GCST90027158)

[4] and used Mendelian randomization (MR) to further investigate the relationship between AD and family history of AD (instrumental variables demonstrated in Additional file 2: Table S1, Additional file 3: Table S2). The results showed a significant bidirectional promoting causal relationship between them (Fig. 2A, Additional file 1: Figs. S1, S2). We suspected that both are driven by the same genetic variants and therefore conducted co-localization analyses in two genomic regions, including the regions near the lead SNP for ebi-a-GCST005921 and near PVRIG genetic locus (a risk gene for AD identified by the author). Within both gene regions, we discovered a very high posterior probability (100% and 99.14%) supporting Hypothesis 4 (H4), and two co-localized genetic loci (rs117310449 and rs6979218) were identified respectively (Fig. 2B, C, Table 1). Conclusively, to some extent, the family history of AD may be able to be used as a substitute for the onset of AD, but there are significant limitations that need to be discussed in the study. In addition, a large amount of GWAS datasets on AD disease have been shared in several public databases (IEU Open GWAS, GWAS Catalog), so there is no need to investigate the relationship between AD and cancer by using GWAS data on family history of AD. We suggest that the authors replace the research question in the paper with the relationship between AD family history and cancer, which is a very research-valuable question as well; and the relationship between AD and cancer needs to be further researched with the correct dataset.

[†]Jia-Cheng Xiang and Yi-Fan Xiong share co-first authorship and contributed equally to this work.

[†]Shao-Gang Wang and Qi-Dong Xia contributed equally to this work.

*Correspondence:

Shao-Gang Wang
sgwangtjm@163.com

Qi-Dong Xia
qidongxia_md@163.com

¹ Department and Institute of Urology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, No.1095 Jiefang Avenue, Wuhan 430030, China



A

Family history of Alzheimer's disease

Dataset: ebi-a-GCST005921

Download VCF Download index	
PMID	29777097
Year	2018
Category	NA
Sub category	NA
Population	European
Sex	NA
Sample size	314,278
Number of SNPs	7,746,640
Unit	NA
Author	Marioni RE
Consortium	NA
Ontology	NA
Build	HG19/GRCh37
Note	NA

B

Ever had bowel cancer screening

Dataset: ukb-b-17001

Download VCF Download index View report	
Year	2018
Category	Binary
Sub category	NA
Population	European
Sex	Males and Females
ncase	146,163
ncontrol	309,096
Sample size	455,259
Number of SNPs	9,851,867
Unit	SD
Priority	1
Author	Ben Elsworth
Consortium	MRC-IEU
Ontology	NA

C

Ever had bowel cancer screening

Dataset: ukb-a-296

Download VCF Download index View report	
Year	2017
Category	NA
Sub category	NA
Population	European
Sex	Males and Females
ncase	107,072
ncontrol	224,939
Sample size	332,011
Number of SNPs	10,894,596
Unit	SD
Priority	1
Author	Neale
Consortium	Neale Lab
Ontology	NA

Fig. 1 IEU Open GWAS search results for **A** ebi-a-GCST005921, **B** ukb-b-17001, and **C** ukb-a-296

Additionally, the authors extracted eQTL data of brain tissue and whole blood from the GTEX database and identified PVRIG as a risk gene for AD by co-localization analysis. We believe that the robustness of the proof process in this section needs to be improved: firstly, the authors performed the analysis using the Coloc R package and the web tool Sherlock, but only reported the

log Bayes factor (LBF) without the posterior probabilities of each hypothesis for the co-localization analysis; secondly, co-localization analysis is mainly adopted to evaluate whether two traits are driven by the same genetic locus, which is insufficient to establish a causal link between them [5], whereas Mendelian randomization can establish a valid causal relationship, however,

A

Exposure	Outcome	Method	Heterogeneity	Pleiotropy	Pval	OR (95% CI)
AD family history (8 SNPs) (ebi-a-GCST005921)	Alzheimer's disease (ebi-a-GCST90027158)	Inverse variance weighted	1.51e-15		4.50e-22*	3.83 (2.92 to 5.03)
		Inverse variance weighted (MRE)			4.50e-22*	3.83 (2.92 to 5.03)
		MR Egger	3.21e-13	0.314	6.10e-02	2.53 (1.15 to 5.57)
		Weighted median			5.47e-27*	3.11 (2.53 to 3.82)
		Simple mode			6.26e-04*	4.27 (2.63 to 6.94)
Alzheimer's disease (43 SNPs) (ebi-a-GCST90027158)	AD family history (ebi-a-GCST005921)	Weighted mode			8.98e-06*	2.99 (2.48 to 3.62)
		Inverse variance weighted	1.68e-08		2.38e-62*	1.54 (1.46 to 1.61)
		Inverse variance weighted (MRE)			2.38e-62*	1.54 (1.46 to 1.61)
		MR Egger	6.68e-08	0.144	2.07e-07*	1.42 (1.27 to 1.59)
		Weighted median			7.63e-40*	1.54 (1.44 to 1.64)
Simple mode			3.67e-04*	1.43 (1.19 to 1.71)		
Weighted mode				2.45e-03*	1.35 (1.13 to 1.63)	

*: P<0.05

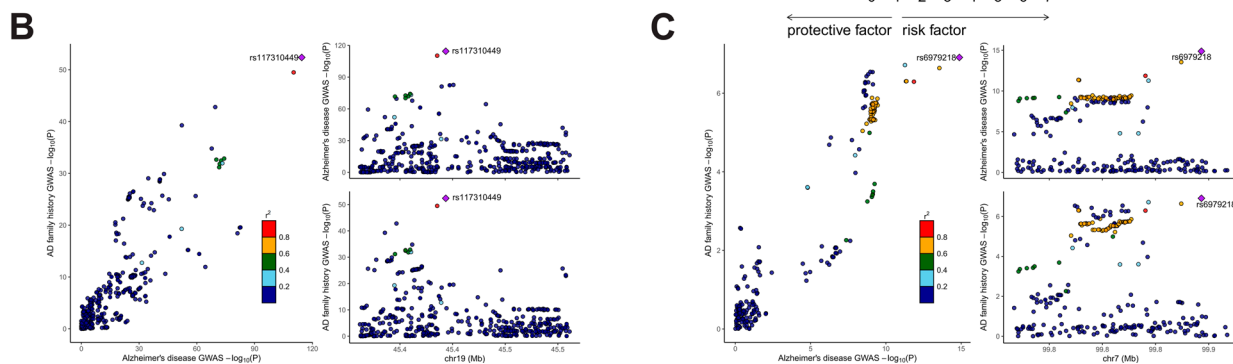


Fig. 2 **A** Results of bidirectional Mendelian randomization between family history of AD and AD; **B** results of co-localization between AD family history and AD in the gene region near lead SNP rs429358 ($\pm 100,000$ bp) of AD family history GWAS data; **C** results of co-localization between AD family history and AD in the region near the PVRIG gene ($\pm 100,000$ bp)

the authors identified PVRIG as a risk gene for AD only after co-localization; finally, the authors used only eQTL data from GTEX without external validation, thus the conclusions remain highly limited to some extent. Therefore, collecting the cis-eQTLs near the PVRIG gene from the eQTLGen database as the exposure (instrumental variables demonstrated in Additional file 4: Table S3), we performed Mendelian randomization analyses to explore the causal relationship between PVRIG and the two AD related traits. Interestingly, the MR results showed that PVRIG was a significant protective factor for both of the AD family history and AD (Fig. 3A, Additional file 1: Figs. S3, S4), contrary to the conclusions obtained by the authors. Reverse MR analysis have ruled out the existence of a reverse causal effect (Additional file 1: Fig. S5). We recommend that the authors perform MR analyses with data from the GTEX database as well. Furthermore, we performed co-localization analyses between PVRIG and

the two AD related traits in the gene region near the lead SNP for PVRIG eQTL data. The results showed that the posterior probability supporting H4 was 80.46% between PVRIG and family history of AD, and 71.21% between PVRIG and AD; the co-localized SNPs were rs705867 and rs55796551, respectively (Fig. 3B, C, Table 1).

In conclusion, we identified some errors in the GWAS datasets used by the authors, which suggests that some of the conclusions have limitation and inaccuracy that require more attention; furthermore, we provided suggestions for the authors to improve analytical methodology and conducted some complementary analyses using data from other sources, which led to some opposite conclusions. Figure 4 summarized our complementary analyses. Although the conclusions of our analyses differ from part of the authors', both of us identify a strong association between PVRIG and AD, and the cellular and molecular mechanisms between them deserve to be further investigated.

Table 1 Results of the posterior probability obtained from co-localization analyses

Exposure	Outcome	nSNPs	Chr	Position	PP.H0.abf	PP.H1.abf	PP.H2.abf	PP.H3.abf	PP.H4.abf	Colocalized SNP
AD family history (ebi-a-GCST005921)	Alzheimer's disease (ebi-a-GCST90027158)	473	19	45411941 ± 100000 bp (around lead SNP: rs429358)	0.00%	0.00%	0.00%	0.00%	100.00%	rs117310449
PVRIG (eqtl-a-ENSG00000213413)	AD family history (ebi-a-GCST005921)	247	7	99716871–99919113 bp (around PVRIG gene loci)	0.00%	0.04%	0.00%	0.81%	99.14%	rs6979218
	AD family history (ebi-a-GCST005921)	301	7	99719480 ± 100000 bp (around lead SNP: rs60458236)	0.00%	0.00%	0.57%	18.97%	80.46%	rs705867
	Alzheimer's disease (ebi-a-GCST90027158)	396	7	99719480 ± 100000 bp (around lead SNP: rs60458236)	0.00%	0.00%	0.00%	28.79%	71.21%	rs55796551

"PP.H0-4.abf" represents the posterior probability supporting the H0-4 hypotheses in co-localization analysis

A

Exposure	Outcome	Method	Heterogeneity	Pleiotropy	Pval	OR (95% CI)
PVRIG (15 SNPs) (eQTL-a-ENSG00000213413)	AD family history (ebi-a-GCST005921)	Inverse variance weighted	1.35e-06		1.64e-04*	0.93 (0.90 to 0.97)
		Inverse variance weighted (MRE)			1.64e-04*	0.93 (0.90 to 0.97)
		MR Egger	1.10e-06	0.568	3.76e-01	0.96 (0.87 to 1.05)
		Weighted median			1.19e-07*	0.93 (0.90 to 0.95)
		Simple mode			1.57e-03*	0.92 (0.88 to 0.96)
PVRIG (16 SNPs) (eQTL-a-ENSG00000213413)	Alzheimer's disease (ebi-a-GCST90027158)	Inverse variance weighted	4.79e-09		1.48e-06*	0.89 (0.85 to 0.93)
		Inverse variance weighted (MRE)			1.48e-06*	0.89 (0.85 to 0.93)
		MR Egger	6.04e-08	0.194	3.89e-01	0.95 (0.86 to 1.06)
		Weighted median			2.59e-15*	0.88 (0.86 to 0.91)
		Simple mode			9.63e-04*	0.87 (0.82 to 0.93)
		Weighted mode			1.43e-06*	0.89 (0.86 to 0.91)

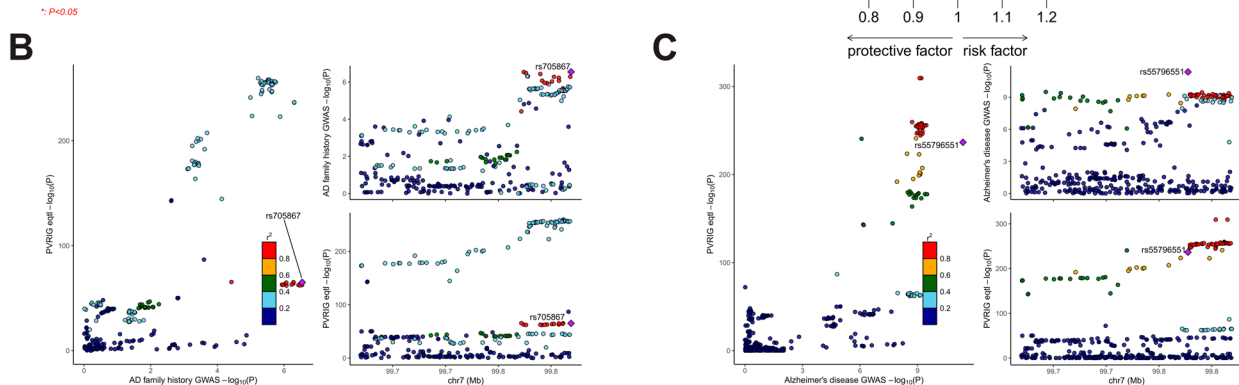


Fig. 3 **A** Mendelian randomization results with PVRIG as exposure and family history of AD and AD as outcome; **B** results of co-localization between PVRIG and AD family history in the gene region near lead SNP rs60458236 ($\pm 100,000$ bp) of PVRIG eQTL data; **C** results of co-localization between PVRIG and AD in the gene region near lead SNP rs60458236 ($\pm 100,000$ bp) of PVRIG eQTL data

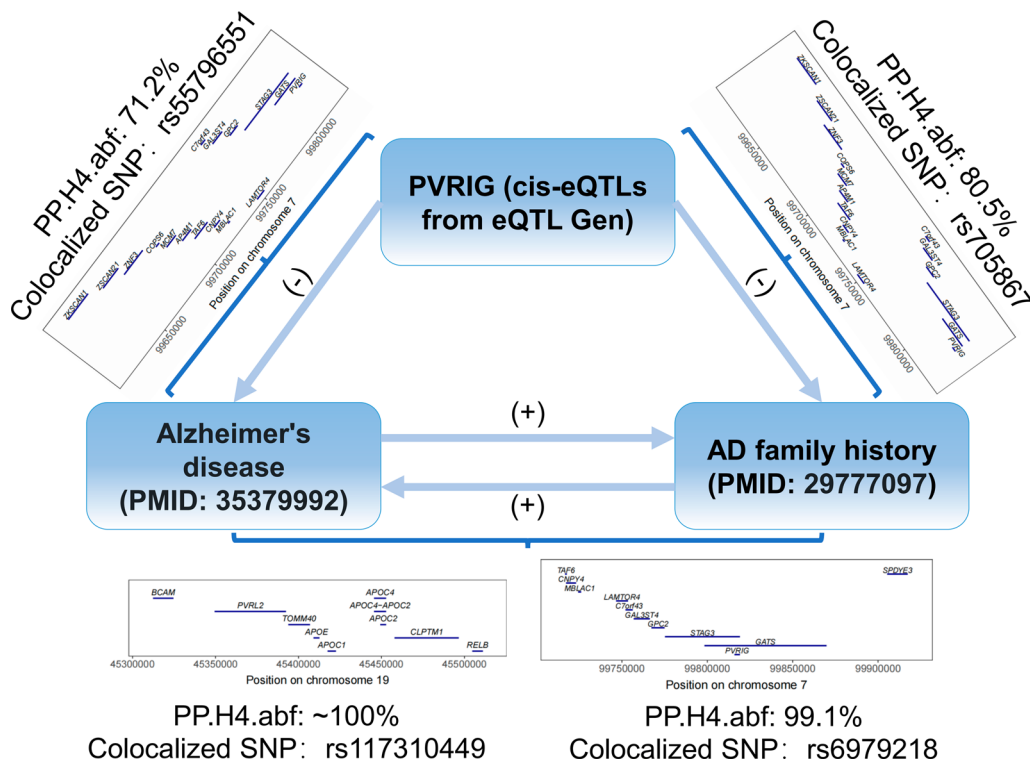


Fig. 4 Summary of the MR analysis and the results of the co-localization analysis in this paper

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12967-024-05106-w>.

Additional file 1: Figure S1. Mendelian randomization with family history of AD as the exposure and AD as the outcome, this figure showed (A) scatterplot, (B) leave-one-out test plot, (C) funnel plot, and (D) forest plot, respectively. **Figure S2.** Mendelian randomization with AD as the exposure and family history of AD as the outcome, this figure showed (A) scatterplot, (B) leave-one-out test plot, (C) funnel plot, and (D) forest plot, respectively. **Figure S3.** Mendelian randomization with PVRIG as the exposure and family history of AD as the outcome, this figure showed (A) scatterplot, (B) leave-one-out test plot, (C) funnel plot, and (D) forest plot, respectively. **Figure S4.** Mendelian randomization with PVRIG as the exposure and AD as the outcome, this figure showed (A) scatterplot, (B) leave-one-out test plot, (C) funnel plot, and (D) forest plot, respectively. **Figure S5.** Mendelian randomization results with family history of AD and AD as exposure and PVRIG as outcome.

Additional file 2: Table S1. Instrumental variables selected from "ebi-a-GCST005921" for AD family history. Filtering condition: $P < 5e-8$; Clump: $kb = 10000$, $r^2 = 0.001$.

Additional file 3: Table S2. Instrumental variables selected from "ebi-a-GCST90027158" for AD. Filtering condition: $P < 5e-8$; Clump: $kb = 10000$, $r^2 = 0.001$.

Additional file 4: Table S3. Instrumental variables selected from "eqtl-a-ENSG00000213413" for PVRIG. Filtering condition: $P < 5e-8$; Clump: $kb = 100$, $r^2 = 0.3$.

Acknowledgements

We thank xiechengyong123 for the acquisition and instruction of the R package "friendly2MR" (<https://github.com/xiechengyong123/friendly2MR>).

Author contributions

XJC, XQD and WSG contributed to study concept and design. XJC and XYF contributed to acquisition of data. XJC and XYF performed the Mendelian randomization analysis. XJC visualized the data and wrote the first draft of the manuscript. XQD and WSG revised the manuscript. All authors contributed to the article and approved the submitted version. Jia-Cheng Xiang; Yi-Fan Xiong; Qi-Dong Xia; and Shao-Gang Wang contributed equally to this work.

Funding

Not applicable.

Availability of data and materials

The original contributions presented in the study are included in the article/ supplementary material, further inquiries can be directed to the corresponding authors.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

We declared that there is no conflict of interest.

References

- Zheng H, Shi Y-Z, Liang J-T, Lu L-L, Chen M. Modifiable factors for migraine prophylaxis: a Mendelian randomization analysis. *Front Pharmacol*. 2023;14:1010996.
- Marioni RE, Harris SE, Zhang Q, McRae AF, Hagenaars SP, Hill WD, et al. GWAS on family history of Alzheimer's disease. *Transl Psychiatry*. 2018;8(1):99.
- Gatz M, Reynolds CA, Fratiglioni L, Johansson B, Mortimer JA, Berg S, et al. Role of genes and environments for explaining Alzheimer disease. *Arch Gen Psychiatry*. 2006;63(2):168–74.
- Bellenguez C, Küçükali F, Jansen IE, Kleindam L, Moreno-Grau S, Amin N, et al. New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nat Genet*. 2022;54(4):412–36.
- Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet*. 2014;10(5): e1004383.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.