

Genomic variation in educational attainment modifies Alzheimer disease risk

Neha S. Raghavan, PhD, Badri Vardarajan, PhD, and Richard Mayeux, MD

Neurol Genet 2019;5:e310. doi:10.1212/NXG.0000000000000310

Correspondence

Dr. Mayeux
rpm2@columbia.edu

Abstract

Objective

To determine the putative protective relationship of educational attainment on Alzheimer disease (AD) risk using Mendelian randomization and to test the hypothesis that by using genetic regions surrounding individually associated single nucleotide polymorphisms (SNPs) as the instrumental variable, we can identify genes that contribute to the relationship.

Methods

We performed Mendelian randomization using genome-wide association study summary statistics from studies of educational attainment and AD in two stages. Our instrumental variable comprised (1) 1,271 SNPs significantly associated with educational attainment and (2) individual 2-Mb regions surrounding the genome-wide significant SNPs.

Results

A causal inverse relationship between educational attainment and AD was identified by the 1,271 SNPs (odds ratio = 0.63; 95% confidence interval, 0.54–0.74; $p = 4.08 \times 10^{-8}$). Analysis of individual loci identified 2 regions that significantly replicated the causal relationship. Genes within these regions included *LRRC2*, *SSBP2*, and *NEGR1*; the latter a regulator of neuronal growth.

Conclusions

Educational attainment is an important protective factor for AD. Genomic regions that significantly paralleled the overall causal relationship contain genes expressed in neurons or involved in the regulation of neuronal development.

From the The Gertrude H. Sergievsky Center (N.S.R., B.V.), Columbia University; The Institute for Genomic Medicine (N.S.R.), Columbia University; The Taub Institute for Research in Alzheimer's Disease and the Aging Brain (B.V.), Columbia University; The Department of Neurology (R.M.), Columbia University and The New York Presbyterian Hospital; The Department of Epidemiology (R.M.), Joseph P. Mailman School of Public Health, Columbia University, New York, NY.

Funding information and disclosures are provided at the end of the article. Full disclosure form information provided by the authors is available with the full text of this article at Neurology.org/NG.

The Article Processing Charge was funded by the authors.

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND), which permits downloading and sharing the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

Glossary

AD = Alzheimer disease; **CI** = confidence interval; **GWAS** = genome-wide association study; **IGAP** = International Genomics of Alzheimer's Project; **IVW** = inverse variance weighted; **LD** = linkage disequilibrium; **OR** = odds ratio; **SNP** = single nucleotide polymorphism.

Education has consistently been identified as an important antecedent factor in Alzheimer disease (AD), whereby advanced educational attainment is thought to reduce AD risk.^{1,2} In addition, educational attainment and AD may be genetically related based on the genome-wide correlation of -0.31 ($p = 4 \times 10^{-4}$).³ However, confounding factors that affect educational attainment such as socioeconomic status, nutrition, and ethnicity blur the relationship.

Mendelian randomization limits confounding by using instrumental variables associated with a risk factor to establish effects on the outcome. The largest ($n = 1,131,881$) genome-wide association study (GWAS) of educational attainment⁴ identified 1,271 significantly associated single nucleotide polymorphisms (SNPs), and a separate 2013 GWAS from the International Genomics of Alzheimer's Project (IGAP) provided AD SNP association data. The data created an opportunity to examine the role of educational attainment in AD through Mendelian randomization.

In GWAS analyses, the significantly associated tag SNPs may not be the causal SNP.⁵ When pleiotropic effects and heterogeneity remain minimal, the instrumental variable composed of the index and all associated SNPs in a region improves the validity of the relationship,⁶ explains a greater percentage of variation in the phenotype, and reduces heterogeneous effects of using fewer SNPs.⁷

In addition to testing the overall hypothesis that educational attainment has a protective effect on AD risk, we tested the hypothesis that a subset of genetic loci related to educational attainment might individually significantly reflect the overall protective relationship with AD. To maximize detection, we included independent SNPs in a 2-Mb region surrounding each of the 1,271 SNPs significantly associated with educational attainment.³

Methods

We used available summary statistics from the analysis of all discovery data excluding the 23andMe cohort in the largest GWAS meta-analysis of educational attainment³ ($n = 766,345$). Summary statistics from the IGAP GWAS of AD,⁸ consisting of 54,162 individuals and 7,055,881 analyzed SNPs, were used for the outcome. There is a small overlap of samples used from the educational attainment and AD GWAS. Sample overlap in a 2-sample Mendelian randomization can cause bias in results; however, because of the large sample size of the educational attainment GWAS and minimal overlap of the sample,

no appreciable bias was expected. The study population in both GWASs were of European descent.

SNP selection

TwoSampleMR, an R package that performs Mendelian randomization using data from MR-Base,⁹ was used in R (R v.3.3.1) to perform all analyses. Mendelian randomization was performed using two different schemes for selecting SNPs for the instrumental variable:

- 1,271 SNP analysis: the instrumental variables consisted of the 1,271 approximately independent SNPs (all with $p < 5 \times 10^{-8}$) significantly associated with educational attainment and were used to test for causality to AD. Independence for SNPs in the analysis was established using linkage disequilibrium (LD) clumping ($r^2 \geq 0.001$ within a 10,000-kb window).
- Individual locus analyses: Mendelian randomization was performed independently on each of the 1,271 loci using SNPs within 1-Mb upstream and downstream of individual SNPs. We used more liberal inclusion criteria including SNPs associated with educational attainment at $p < 0.01$ and clumped at $r^2 \geq 0.1$ within a 250-kb window.

Of the 1,271 independent loci, SNPs found within a 2-Mb region of another were merged together, resulting in 441 independent loci.

Mendelian randomization analyses

The strictest LD threshold was applied for the first analysis, using default settings in MR-Base; for analysis 2, we allowed for more SNPs with still minimal LD to be included in the analysis, as we hypothesized that this would strengthen the per-loci analysis. Those SNPs remaining after LD clumping were queried within the IGAP summary statistics, if they were not present in that data set, genetic proxies were found, and finally SNPs for which no proxies could be found were excluded. Next, SNPs were harmonized for the effect allele between the 2 GWAS data sets or removed if harmonization predictions were inconclusive. For the joint 1,271 SNP analysis, 387 remained after LD clumping, and 13 SNPs were neither found in the IGAP nor had appropriate proxies and thus removed. Finally, 57 SNPs were removed because of low confidence that the effect allele of the exposure corresponded to the same allele in the outcome. Therefore, a total of 317 SNPs were included in the initial joint SNP analysis. For the regional analyses, SNPs were similarly removed if not found in the IGAP or if allele harmonization failed. The inverse variance weighted (IVW) method was used to calculate odds ratios (ORs). Results are reported as the OR ($\pm 95\%$ confidence

Table Results from inverse variance weighted method and sensitivity analyses for Mendelian randomization analyses

				Inverse variance weighted method	Sensitivity analysis			
Top educational attainment SNPs (317 SNPs included)				Odds ratio = 0.63; 95% CI, 0.54–0.74; $p = 4.08 \times 10^{-8}$	Egger regression: intercept = 0.003; p value = 0.43			
Locus ^a	Chromosome	Gene(s) ^b	Proposed role of gene(s)	No. of SNPs ^c	Inverse variance weighted method		Sensitivity analysis	
					Odds ratio (95% CI)	p Value	Egger regression (intercept \pm SE, p value)	
rs7552964, rs10789285, rs1024268, rs663251, rs481940, rs72677177, rs34122915, rs34305371, rs2568955, rs1445591, rs12028229, rs11210228, rs74091672, rs11210400, rs1569092, rs28482086	1	<i>LRRC7/PTGER3/NEGR1/FPGT-TNNI3K/TNNI3K/TYW3</i>	Necessary for synaptic spine architecture and function/1 receptor for prostaglandin E2/cell adhesion and/or regenerative axon sprouting/read-through transcription/MAPKKK family involved in cardiac physiology/stabilizes codon-anticodon interactions	89	0.31(0.19–0.50)	1.92×10^{-6}	0.013 ± 0.006 , p value = 0.05	
rs1910005	5	<i>SSBP2</i>	DNA damage response; maintenance of genome stability; telomere repair	17	0.07 (0.02–0.25)	5.8×10^{-5}	-0.008 ± 0.017 , p value = 0.66	

Abbreviation: SNP = single nucleotide polymorphism.

Top: 1,271 SNP Analysis.

Bottom: Individual locus analyses.

^a SNPs significantly associated with educational attainment in Lee et al.,⁴ in this locus.

^b The gene or closest gene to the SNP.

^c Number of SNPs that were included in the regional analysis.

interval [CI]) of AD risk per SD increase in educational attainment in each test.

Sensitivity analyses

Mendelian randomization requires that the instrumental variables meet 3 requirements: it must be associated with the risk factor, not associated with any confounder of the risk factor or outcome, and is only associated with the outcome through the risk factor. Sensitivity analyses following the initial analysis provide confidence that assumptions of the instrumental variables were not broken. In addition to the IVW, the weighted median method was used to measure causality and provided consistent results with IVW when at least 50% of the instrumental variables were valid. Next, the intercept of the Mendelian randomization–Egger test was used to determine potential horizontal pleiotropy or an effect of the instrumental variables on a phenotype other than the outcome. As the intercept neared zero in the mendelian randomization–Egger test, horizontal pleiotropy was reduced. The Steiger test of directionality was used to confirm directionality of the effect, i.e., that the SNPs first affected educational attainment and subsequently that AD risk was affected through educational attainment. Finally, the radial regression analyses were performed for the inverse variance methods, which identify any significant SNP outliers using the Cochran’s Q-statistic.¹⁰

Results

Using the 1,271 SNPs associated with educational attainment, we detected statistically significant evidence for causation between educational attainment and AD such that an increase of 4.2 years of educational attainment was associated with a 37% reduction in AD risk (OR—scaled per SD, 4.2 years = 0.63; 95% CI, 0.54–0.74; $p = 4.08 \times 10^{-8}$).

In the per-loci analyses of the 1,271 regions, 2 independent SNP regions demonstrated a statistically significant inverse relationship between educational attainment and AD risk (Bonferroni-corrected threshold $p < 1.2 \times 10^{-4}$). These regions include the neuronal growth regulator precursor (*NEGR1*) gene, leucine-rich repeat containing 7 (*LRCC7*) gene, and prostaglandin E receptor 3 (*PTGER3*) gene (table 1).

Sensitivity analyses were performed for all Mendelian randomization analyses. There was no indication of pleiotropy, reverse causality, or heterogeneity, and the weighted median method showed consistent results with IVW overall in the significant analyses.

Discussion

Consistent with earlier reports, we found an inverse relationship of educational attainment with AD.^{11–13} In addition, we also identified regions that individually significantly replicated the causal relationship of education on AD, several of which were found to contain genes expressed in neurons or

involved in regulation of neuronal development. For example, one of the SNPs is within the *LRCC7* gene, which is crucial to dendritic spine architecture and function and may be involved in bipolar disorder.¹⁴ Another SNP within the intronic region of the *NEGR1* gene is also highly expressed in neurons and has been found to be associated with major depressive disorder.¹⁵ An SNP was found in the *PTGER3* gene, which is thought to be involved in the modulation of neurotransmitter release in neurons.

One limitation of using the Mendelian randomization method for outcomes with other strong nongenetic factors risk factors is that the CIs tend to be large. However, this outcome is preferable to the bias that is inherent in epidemiologic studies, which unavoidably include confounders.¹⁶ Taken together, the results presented here, along with earlier reports, establish a putative causal relationship between educational attainment and AD.

The key next steps are to replicate these findings in diverse ethnic groups with more variable educational experiences and to identify and validate specific variants within these loci that account for the association.

Author contributions

All authors of the study contributed to the study design, study analysis, and writing and editing.

Acknowledgments

The authors thank the International Genomics of Alzheimer’s Project and Lee et al. for providing summary statistics for this project.

Study funding

This work was supported by grants from the National Institute on Aging from the NIH (U01AG032984 and RO1AG041797) and the National Center for Advancing Translational Sciences (TL1TR001875).

Disclosure

N.S. Raghavan reports no disclosures. B. Vardarajan has served as a consultant for BISC Global. R. Mayeux has served as a consultant for the Alzheimer’s Disease Center at Rush University Medical School and has received research support from the National Institute on Aging. Full disclosure form information provided by the authors is available with the full text of this article at Neurology.org/NG.

Publication history

Received by *Neurology: Genetics* July 15, 2018. Accepted in final form December 18, 2018.

References

1. Caamano-Isorna F, Corral M, Montes-Martinez A, Takkouche B. Education and dementia: a meta-analytic study. *Neuroepidemiology* 2006;26:226–232.
2. Barnes DE, Yaffe K. The projected effect of risk factor reduction on Alzheimer’s disease prevalence. *Lancet Neurol* 2011;10:819–828.
3. Okbay A, Beauchamp JP, Fontana MA, et al. Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* 2016;533:539–542.

4. Lee JJ, Wedow R, Okbay A, et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat Genet* 2018;50:1112–1121.
5. Schaid DJ, Chen W, Larson NB. From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat Rev Genet* 2018;19:491–504.
6. Burgess S, Bowden J, Fall T, Ingelsson E, Thompson SG. Sensitivity analyses for Robust causal inference from mendelian randomization analyses with multiple genetic variants. *Epidemiology* 2017;28:30–42.
7. Pierce BL, Ahsan H, Vanderweele TJ. Power and instrument strength requirements for Mendelian randomization studies using multiple genetic variants. *Int J Epidemiol* 2011;40:740–752.
8. Lambert JC, Ibrahim-Verbaas CA, Harold D, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet* 2013;45:1452–1458.
9. Hemani G, Zheng J, Wade KH, et al. MR-base: a platform for systematic causal inference across the phenome using billions of genetic associations. *bioRxiv* 2016.
10. Bowden J, Spiller W, Del Greco MF, et al. Improving the visualization, interpretation and analysis of two-sample summary data Mendelian randomization via the radial plot and radial regression. *Int J Epidemiol* 2018;47:1264–1278.
11. Anderson E, Wade KH, Hemani G, et al. The causal effect of educational attainment on Alzheimer's disease: a two-sample Mendelian randomization study. *bioRxiv* 2017.
12. Larsson SC, Traylor M, Malik R, Dichgans M, Burgess S, Markus HS. Modifiable pathways in Alzheimer's disease: mendelian randomisation analysis. *BMJ* 2017;359:j5375.
13. Østergaard SD, Mukherjee S, Sharp SJ, et al. Associations between potentially modifiable risk factors and Alzheimer disease: a mendelian randomization study. *PLoS Med* 2015;12:e1001841.
14. Fiorentino A, Sharp SJ, Kandaswamy R, Gurling HM, Bass NJ, McQuillin A. Genetic variant analysis of the putative regulatory regions of the LRRC7 gene in bipolar disorder. *Psychiatr Genet* 2016;26:99–100.
15. Hyde CL, Nagle MW, Tian C, et al. Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. *Nat Genet* 2016;48:1031–1036.
16. Bautista LE, Smeeth L, Hingorani AD, Casas JP. Estimation of bias in non-genetic observational studies using mendelian triangulation. *Ann Epidemiol* 2006;16:675–680.