


OPEN

Evidence of a Causal Association Between Cancer and Alzheimer's Disease: a Mendelian Randomization Analysis

Sahba Seddighi^{1,2}, Alexander L. Houck³, James B. Rowe⁴ & Paul D. P. Pharoah⁵ 

While limited observational evidence suggests that cancer survivors have a decreased risk of developing Alzheimer's disease (AD), and vice versa, it is not clear whether this relationship is causal. Using a Mendelian randomization approach that provides evidence of causality, we found that genetically predicted lung cancer (OR 0.91, 95% CI 0.84–0.99, $p = 0.019$), leukemia (OR 0.98, 95% CI 0.96–0.995, $p = 0.012$), and breast cancer (OR 0.94, 95% CI 0.89–0.99, $p = 0.028$) were associated with 9.0%, 2.4%, and 5.9% lower odds of AD, respectively, per 1-unit higher log odds of cancer. When genetic predictors of all cancers were pooled, cancer was associated with 2.5% lower odds of AD (OR 0.98, 95% CI 0.96–0.988, $p = 0.00027$) per 1-unit higher log odds of cancer. Finally, genetically predicted smoking-related cancers showed a more robust inverse association with AD than non-smoking related cancers (OR 0.95, 95% CI 0.92–0.98, $p = 0.0026$, vs. OR 0.98, 95% CI 0.97–0.995, $p = 0.0091$).

Individuals with a history of cancer are less likely to develop Alzheimer's disease (AD), and vice versa^{1–3}. This epidemiological observation remains unexplained, although biologically plausible mechanisms have been proposed: while cancer is a disorder of excessive cell proliferation, neurodegeneration is one of premature cell death. The dysregulation of mutual genes, proteins, and pathways in both conditions has been demonstrated⁴. For instance, the expression of the tumor suppressor TP53, which is notably downregulated in the majority of cancers, is upregulated in AD. The opposite is true for PIN1, an enzyme promoter of cell proliferation⁴.

The magnitude of the association between AD and cancer varies greatly across studies. In one population-based prospective study, those with prevalent cancer had a 69% lower risk of developing AD after controlling for numerous factors⁵. Another found that those with a history of cancer had just a 13% lower risk of subsequent AD⁶. Others report a trend without reaching statistical significance^{7,8}, or no association between a history of cancer and the risk of AD⁹. In light of these findings, it is important to consider the potential for bias and confounding behind these associations. The reduced risk of AD in those diagnosed with cancer may be due to higher mortality rates among cancer patients and survivors, especially for cancers with poor survival rates, such as pancreatic cancer¹⁰. Many studies are limited by small sample sizes^{7,11,12}, short follow-up periods^{7,9,12,13}, or gender bias⁹. Other issues include appropriate diagnoses and changing guidelines, the potential negative effects of chemotherapy on cognition, and behavioral and sociodemographic variables related to both AD and cancer.

A fundamental question remains: is cancer causally related to AD, or is the observed epidemiological relationship between the two disorders an artifact of study design, confounding and biases? This study uses a genetic approach to determine the nature of the association between cancer and AD. The alleles of cancer-associated genetic variants are randomly allocated and are not typically associated with the range of social, behavioral, and physiological factors that can influence the observed relationship between cancer and Alzheimer's disease. Thus, the method of Mendelian randomization, in which genetic variants are used as a proxy for cancer, represents a valid alternative to assess the causal relationship between cancer and Alzheimer's disease by avoiding issues of confounding or reverse-causality¹⁴.

¹University of Cambridge Institute of Public Health, University of Cambridge, Cambridge, UK. ²Johns Hopkins University School of Medicine, Baltimore, USA. ³University of Tennessee Health Science Center, Memphis, Tennessee, USA. ⁴Department of Clinical Neurosciences, University of Cambridge, Cambridge, UK. ⁵Department of Oncology, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK. Correspondence and requests for materials should be addressed to P.D.P.P. (email: pp10001@medschl.cam.ac.uk)

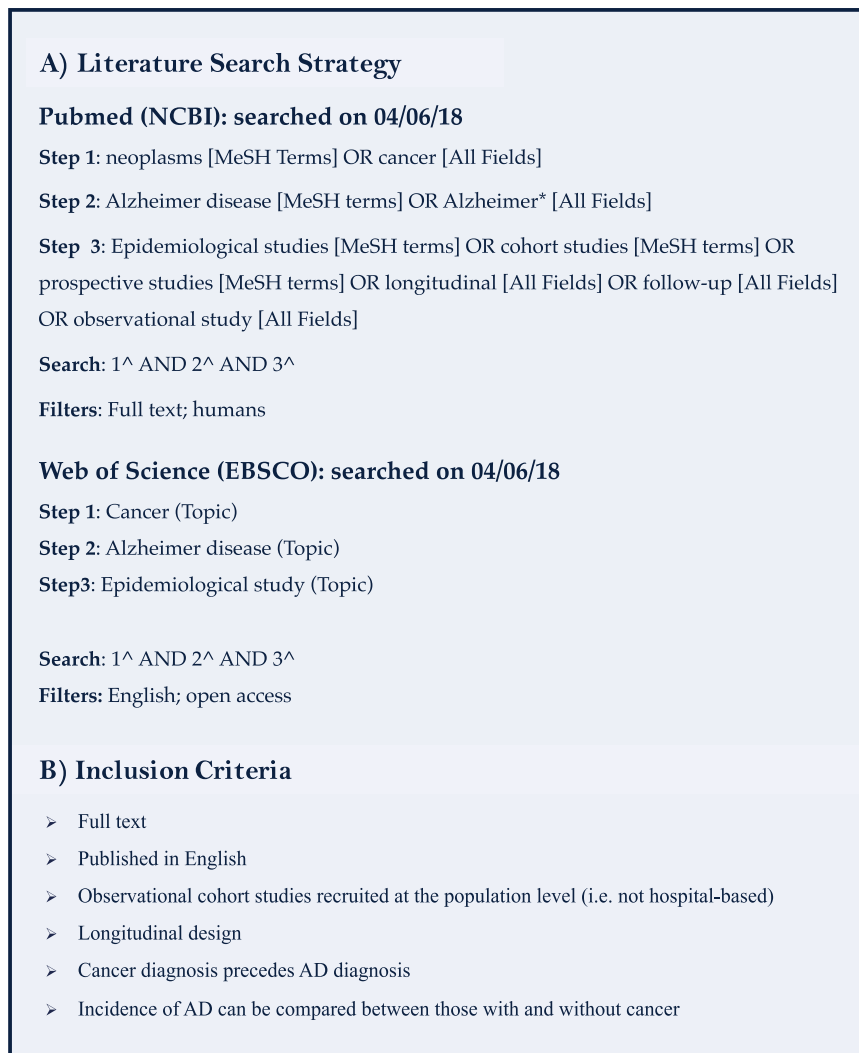


Figure 1. Literature review search strategy.

Prior literature was used to identify genetic variants that influence susceptibility to cancer. Data from the International Genomics of Alzheimer's Project (IGAP) were then utilized to examine the theoretically unfounded association between these cancer-predicting genes and the risk of AD. This approach addresses two major limitations of previous observational studies: the potential for reverse-causation and the confounding effects of environmental risks¹⁵. We predicted that genes that have been associated with increased risk of cancer are associated with reduced risk of AD, in independent cohorts.

Methods

Overview. The summary-data-based Mendelian randomization method consists of three key steps in data preparation and cleaning: (1) the identification of single nucleotide polymorphisms (SNPs) associated with the exposure (in this case, cancer), (2) the determination of SNP-effects on the outcome (in this case, AD), and (3) Mendelian randomization analysis to examine the causal effect of the exposure on the outcome.

Defining instrumental variables for cancer. *Identifying relevant cancer types: systematic literature review.* To identify cancer types with epidemiological evidence of association with AD, PubMed (NCBI) and Web of Science (EBSCO) were systematically searched using a combination of medical subject headings [MeSH] and keyword strings (Fig. 1A). Identified papers were then examined against a predetermined set of inclusion criteria to determine eligibility (Fig. 1B).

Identifying cancer-associated genetic variants: review of GWAS literature. After identifying cancers with prior evidence of association with AD, the Genome-Wide Association Studies (GWAS) Catalog was used to identify genetic variants associated with each cancer type (04/06/18)¹⁶. The GWAS Catalog (<https://www.ebi.ac.uk/gwas/docs/about>) was launched in 2008 by the National Human Genome Research Institute to systematically catalogue and summarize SNP-trait associations from all published GWAS¹⁶. The repository is updated on a weekly basis by an experienced team of molecular biologists with support from The European Bioinformatics

Institute (EMBL-EBI) and input from an independent Scientific Advisory Board. This tool was used to identify published GWAS reporting cancer susceptibility SNPs. Each paper was then manually examined for content and data retrieval.

Only studies conducted or replicated in individuals of European descent were considered in order to enable meaningful comparison with the population used in IGAP for Mendelian randomization outcome data. Furthermore, studies conducted on susceptible subpopulations (e.g. those with a family history of cancer, or ever-smokers) were not considered. SNPs associated with cancer at the standard threshold for genome-wide significance ($p < 5 \times 10^{-8}$) were selected.

A central assumption of Mendelian randomization is that instrumental variables are free of linkage disequilibrium with one another (i.e. each genetic variant must be inherited independently of all others under consideration). Analyses were, therefore, limited to genetic variants that were not in linkage disequilibrium (defined as $r^2 < 0.2$) with other genetic variants for the same type of cancer in order to avoid a violation of this assumption. Pairwise linkage analysis of SNPs was undertaken using the National Cancer Institute LDMatrix tool (<https://analysisstools.nci.nih.gov/LDlink/?tab=ldmatrix>) based on the European subpopulation—Utah Residents from North and West Europe (CEU), Toscani in Italia (TSI), British in England and Scotland (GBI), and Iberian population in Spain (IBS)—reference panel of the 1000 Genomes Project. For genetic variants in linkage disequilibrium, the variant with the lowest p-value for association with cancer was selected. Similarly, for each cancer susceptibility locus, the lead SNP, representing the variant with the lowest p-value of association, was selected.

Identifying proxy SNPs. When cancer-associated genetic variants were unavailable in the outcome dataset, proxies in pairwise linkage disequilibrium (defined by $r^2 > 0.9$) were used, where available. Identification of proxy SNPs was undertaken using the National Cancer Institute LDlink platform's LDproxy tool (<https://analysisstools.nci.nih.gov/LDlink/?tab=ldproxy>) based on the European subpopulation—CEU, TSI, GBI, and IBS—reference panel of the 1000 Genomes Project. Proxy SNP effect alleles were assigned according to correlation information between alleles, provided on the LDpair tool provided through the National Cancer Institute LDlink platform.

Obtaining effects of cancer-associated SNPs on AD. Summary statistics describing the association between each cancer-related SNP and risk of AD were obtained from IGAP^{17,18}. IGAP includes genotyped and imputed data on SNPs from 17,008 AD cases and 37,154 controls of European ancestry from four genome-wide association studies: (1) The Cohorts for Heart and Aging Research in Genomic Epidemiology consortium (CHARGE), (2) The Alzheimer's Disease Genetics Consortium (ADGC), (3) The Genetic and Environmental Risk in Alzheimer's disease consortium (GERAD), and (4) The European Alzheimer's disease Initiative (EADI). A summary of each dataset is provided in Supplementary Table 1. Diagnostic criteria for AD used in each study are provided in Supplementary Table 2.

Imputation and SNP selection. SNPs with call rates $< 95\%$ were excluded from consideration in IGAP. The genotypes of all individuals were imputed with haplotypes from samples of European ancestry in the 1000 Genome Project using IMPUTE2¹⁹ or MaCH/Minimac²⁰. SNPs with a minor allele frequency of $< 1\%$ or with R^2 (MaCH) or info score quality (IMPUTE2) estimates less than 0.3 were excluded from analyses. A total of 8,133,148 SNPs was retained for analysis.

In each dataset, the association between genotype dosage and AD was analyzed by logistic regression. The model was adjusted for age, sex, and principal components to account for possible population stratification. For the CHARGE cohorts containing incident AD cases, Cox proportional hazards regression models were used instead. The consortia used different, but comparable software for these analyses (PLINK (ACT, ADC, ADNI, AGES, GSK, MAYO, OHSU, ROSMAP, TGEN2, UVMUMSS, and UPITT), SNPTEST (GERAD), ProbABEL (AGES, CHS, EADI, FHS, and RS) or R (LOAD, MIRAGE)^{21–23}. SNPs with logistic regression $|\beta| > 5$ or p-value equal to 0 or 1 were excluded. The maximum number of SNPs in any data set was 8,131,643. SNPs that were genotyped/imputed in at least 40% of AD cases and 40% of controls were included in the meta-analysis. This led to a final number of 7,055,881 SNPs for analysis.

Statistical analysis. For the meta-analysis, fixed-effects inverse variance-weighted meta-analysis was applied, with the standard errors of the β -coefficient scaled by the square roots of study-specific genomic inflation factors (estimated before combining summary statistics across all datasets). METAL and GWAMA software packages were used for these analyses^{24,25}.

Statistical analysis of two-sample Mendelian randomization. *Calculating inverse-variance weighted ratio estimates and evaluating the role of chance.* Two-sample Mendelian randomization is a statistical method that can be applied to summary statistics from GWAS to estimate the causal effect of an exposure (in this case, cancer) on an outcome (in this case, AD). “Two-sample” refers to the fact that summary association results for the exposure and the outcome are estimated in non-overlapping sets of individuals²⁶.

An instrumental variable ratio estimate was calculated for each cancer-associated SNP as follows: First, the exposure and outcome files were harmonized on SNPid and effect allele. Then, the effect size estimate (β) for the association of the SNP with risk of AD was divided by the effect size estimate for the association of the same SNP with risk of cancer, producing a ratio of association estimates for each SNP, $(\hat{\theta}_j)$ as follows²⁷:

$$\hat{\theta}_j = \frac{\hat{\beta}_{Yj}}{\hat{\beta}_{Xj}} \quad (1)$$

The ratio estimates from each SNP were then averaged using the inverse-variance weighted formula, adopted from the meta-analysis literature, to produce an overall causal estimate, the inverse-variance weighted (IVW) estimate. First-order weights were selected for the IVW method, as uncertainty in effect estimates was minimized through the use of large sample sets in identifying genetic instruments for cancer. The variance term was calculated as²⁷:

$$\frac{\text{se}(\hat{\beta}_{Yj})^2}{\hat{\beta}_{Xj}^2} \quad (2)$$

And the pooled fixed-effect inverse-variance weighted estimate ($\hat{\theta}_{IVW}$) was calculated as²⁷:

$$\hat{\theta}_{IVW} = \frac{\sum_j \hat{\beta}_{Yj} \hat{\beta}_{Xj} \text{se}(\hat{\beta}_{Yj})^{-2}}{\sum_j \hat{\beta}_{Xj}^2 \text{se}(\hat{\beta}_{Yj})^{-2}} \quad (3)$$

Ratio estimates for each genetic variant were summarized as forest plots. All analyses were conducted in R version 3.5.0 (R foundation) using the two sample Mendelian Randomization package²⁸. The full R script used for two-sample MR is provided in Appendix 1.

To assess the role of chance in our findings, we additionally calculated the Bayesian false-discovery rate (BFDR) for each result, irrespective of statistical significance.

Under frequentist inference, p-values represent the probability of the data given the hypothesis. A Bayesian approach can be used, instead, to assess the probability of a hypothesis given the data. BFDR can be calculated to identify noteworthy associations and prevent overinterpretation of statistically significant findings that are unlikely to be true. This method involves specification of the prior probability that the null hypothesis is false and combines this information with the p-value and study power to derive a more informative posterior probability that the null hypothesis is true. A range of priors was selected on the basis of available observational evidence. All analyses were conducted in R version 3.5.0 (R foundation) using the gap package. The full R script used for BFDR calculations is provided in Appendix 1.

Sensitivity analyses. MR-Egger intercept test: Conventional Mendelian randomization analysis relies on the assumption that genetic variants used as instrumental variables do not have pleiotropic effects—meaning that the chosen variants do not influence the outcome (AD) through any pathways other than the exposure (cancer). This, therefore, implies that the only causal pathway from a cancer-associated genetic variant to AD is via cancer associated biological pathways. While this may be a reasonable assumption when the risk factor under study is a protein biomarker and SNPs are located in or near the coding region for that protein, it is more questionable for polygenic exposures, such as cancer²⁷. We, therefore, used the MR-Egger intercept test to detect violations of this assumption, assessing whether cancer-associated genetic variants have pleiotropic effects on AD that differ on average from zero—known as directional pleiotropy²⁷. If the intercept is not significantly different from 0, there is no evidence to reject the null hypothesis of no directional pleiotropy. MR-Egger intercepts were calculated, using first-order weights, via the two-sample Mendelian randomization package, using R script provided in Appendix 1.

Leave-one-out analysis: To assess the robustness of the Mendelian randomization effect estimates and identify any potential outliers, each SNP was sequentially removed from analysis, conducting Y analyses with Y-1 datapoints²⁷. If the precision and direction of association between cancer-predicting SNPs and AD remain largely unaltered, this implies that the data are not driven by any outliers. Leave-one-out analyses and corresponding plots were generated using the two-sample Mendelian randomization package via R script provided in Appendix 1.

Funnel plots of inverse standard error: Finally, any heterogeneity of effect estimates was visualized using funnel plots depicting the causal effect estimates for each SNP on the x-axis and the inverse standard error (a measure of instrumental strength) for the association on the y-axis. Asymmetry about the vertical line is indicative of heterogeneity; furthermore, a correlation between effect size and instrument strength (which would result in asymmetry) is an indicator of possible directional pleiotropy²⁹. The funnel plots were generated using the two-sample Mendelian randomization package via R script provided in Appendix 1.

Patient involvement: No patients were involved in the design, recruitment, or conduct of this study. No patients were asked to advise on the interpretation of results or the writing of manuscript. Following publication, these results will be available to the general public.

Results

Defining instrumental variables for cancer. *Identifying cancer types of interest.* Following a systematic review of titles, abstracts, and content for eligibility, eleven papers were retrieved from PubMed and Web of Science (Fig. 2)^{5-9,11,13,30-33}. Searching their respective bibliographies did not lead to any additional studies for inclusion. The results of the systematic literature review are summarized in Table 1 and Supplementary Tables 3 and 4.

With the exception of two papers, all studies report an inverse association between cancer and risk of subsequent AD, albeit not always statistically significant at standard thresholds. Nominally significant estimates for a protective effect of cancer on risk of subsequent AD range from 13% reduced risk (HR 0.87, 95% CI 0.84 to 0.90)⁶ to 69% reduced risk (HR 0.31, 95% CI 0.12 to 0.86)⁵. The smallest nominally significant estimate for a protective effect of cancer at any individual site on risk of subsequent AD was seen in individuals with non-melanoma skin cancer (HR 0.95, 95% CI 0.92 to 0.98)³⁰, while the largest estimate was for colorectal cancer (RR 0.43, 95% CI 0.23 to 0.74)¹³. Only data supporting a protective role of cancer in AD reached statistical significance. However,

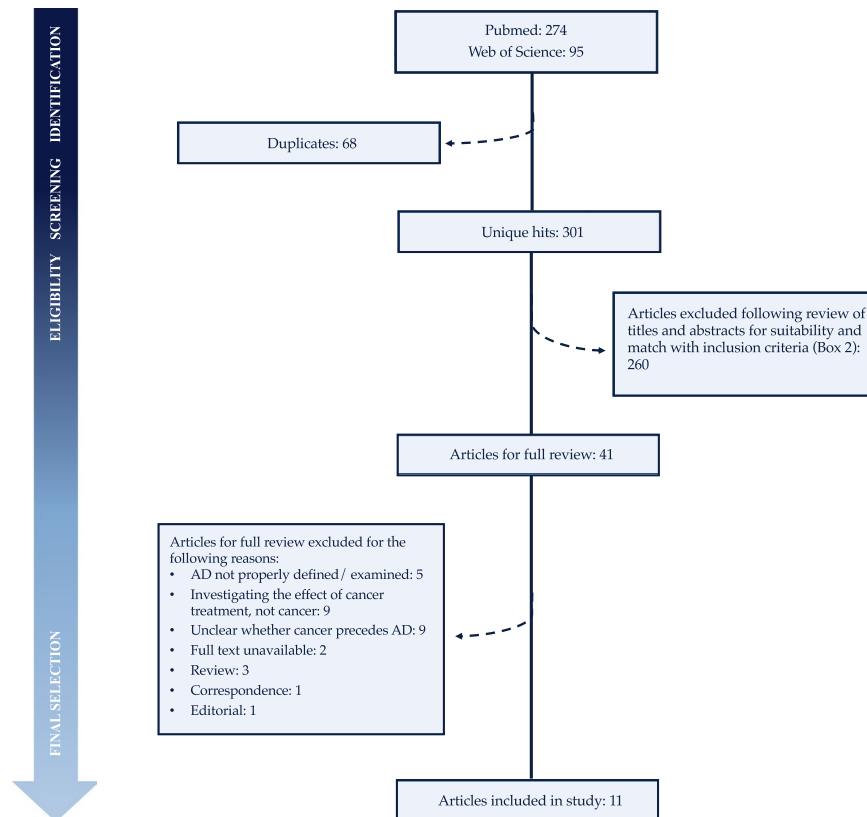


Figure 2. Schematic overview of systematic literature search process and results.

this inverse relationship is not uniformly observed across individual cancer types. For instance, while one study reports a 57% reduction in risk of AD in those with a history of colorectal cancer (RR 0.43, 95% CI 0.23 to 0.74)¹³, another found no nominally significant difference in the risk of AD in those with a history of colorectal cancer (RR 0.97, 95% CI 0.90 to 1.05)⁹. The same is true for non-melanoma skin cancer (HR 0.95, 95% CI 0.92 to 0.98 vs HR 0.18, 95% CI 0.024 to 1.34) (Table 1)^{7,30}.

Furthermore, Driver and colleagues found a more robust inverse association between smoking-related cancers and AD, as compared to non-smoking related cancers and AD (Table 1). In light of these findings, together with the well-established negative association between smoking and Parkinson's disease³⁴, another neurodegenerative disorder, subsequent analyses were stratified by smoking versus non-smoking related cancers to further examine this association in the context of AD. Nominally significant effect estimates from observational studies for smoking-related cancers range from 9% reduced risk of AD (HR 0.91, 95% CI 0.86 to 0.96)⁶ to 40% reduced risk of AD (HR 0.6, 95% CI 0.34 to 0.98)¹³. For non-smoking related cancers, nominally significant effect estimates range from 5% reduced risk of AD (HR 0.95, 95% CI 0.92 to 0.98)³⁰ to 57% reduced risk of AD (RR 0.43, 95% CI 0.23 to 0.74) (Table 1)¹³.

Identifying cancer-associated genetic variants. GWAS data were available for 6 smoking-related cancers—renal cancer carcinoma, cervical cancer, pancreatic cancer, upper aerodigestive tract cancer, urinary bladder cancer, and lung cancer— in the GWAS Catalog repository. Six additional non-smoking related cancers— prostate cancer, leukemia, breast cancer, melanoma, lymphoma, and ovarian cancer— were selected on the basis of SNP information availability within the repository.

Per the inclusion criteria, 60 GWAS were deemed eligible for consideration (Supplementary Table 5). A total of 314 genetic variants were identified as instrumental variables for cancer (Renal carcinoma: 25; pancreatic cancer: 13; upper aerodigestive tract cancer: 14; urinary bladder cancer: 11; lung cancer: 18; cervical cancer: 1; prostate cancer: 36; leukemia: 38; breast cancer: 109; melanoma: 24; lymphoma: 20; ovarian cancer: 19) (Supplementary Tables 5 and 6). The lead SNP (based on the lowest reported p-value from GWAS) at each locus was selected. The lead SNP was also selected for any SNPs in linkage disequilibrium.

Identifying proxy SNPs. Of the 314 identified genetic variants, 26 were not available in IGAP. Seven SNPs had suitable proxies (defined as $r^2 > 0.9$) within the European subpopulation reference panel of the 1000 Genomes Project, and nineteen SNPs did not have suitable proxies (Supplementary Table 7).

Obtaining effects of cancer-associated SNPs on Alzheimer's disease. *International genomics of alzheimer's project.* The International Genomics of Alzheimer's Project datafile consists of the following information for 7,055,881 SNPs and their associations with AD (based on meta-analysis, as described previously):

Cancer type	Reference	Hazard ratio (95% confidence interval)	Covariates
Smoking-related cancers			
Smoking-related**	Driver <i>et al.</i> , 2012 [‡]	0.26 (0.08–0.82)	Age, sex, and smoking
Lung	Musicco <i>et al.</i> , 2013	0.60 (0.34–0.98) [♦]	Age, sex, and calendar year of follow-up
Bladder	Musicco <i>et al.</i> , 2013	0.81 (0.48–1.27) [♦]	Age, sex, and calendar year of follow-up
Smoking-related*	Freedman <i>et al.</i> , 2016	0.91 (0.86–0.96)	Race, sex, and number of doctors' visits, birth year, and cancer registry area
Lung	Frain <i>et al.</i> , 2017	0.79 (0.71–0.89)	Number of clinic visits in the year before baseline, follow-up time, cancer treatment, high cholesterol, hypertension, obesity, coronary arterial disease, diabetes, stroke, and race
Bladder	Frain <i>et al.</i> , 2017	1.01 (0.92–1.11)	Number of clinic visits in the year before baseline, follow-up time, cancer treatment, high cholesterol, hypertension, obesity, coronary arterial disease, diabetes, stroke, and race
Head and neck	Frain <i>et al.</i> , 2017	0.91 (0.80–1.02)	Number of clinic visits in the year before baseline, follow-up time, cancer treatment, high cholesterol, hypertension, obesity, coronary arterial disease, diabetes, stroke, and race
Renal	Frain <i>et al.</i> , 2017	0.79 (0.65–0.97)	Number of clinic visits in the year before baseline, follow-up time, cancer treatment, high cholesterol, hypertension, obesity, coronary arterial disease, diabetes, stroke, and race
Esophagus	Frain <i>et al.</i> , 2017	0.74 (0.52–1.05)	Number of clinic visits in the year before baseline, follow-up time, cancer treatment, high cholesterol, hypertension, obesity, coronary arterial disease, diabetes, stroke, and race
Pancreas	Frain <i>et al.</i> , 2017	0.77 (0.51–1.14)	Number of clinic visits in the year before baseline, follow-up time, cancer treatment, high cholesterol, hypertension, obesity, coronary arterial disease, diabetes, stroke, and race
Stomach	Frain <i>et al.</i> , 2017	0.86 (0.62–1.20)	Number of clinic visits in the year before baseline, follow-up time, cancer treatment, high cholesterol, hypertension, obesity, coronary arterial disease, diabetes, stroke, and race
Non-smoking related cancers			
Non-smoking related	Driver <i>et al.</i> , 2012 [‡]	0.82 (0.57–1.19)	Age, sex, and smoking
Breast	Musicco <i>et al.</i> , 2013	0.70 (0.46–1.03) [♦]	Age, sex, and calendar year of follow-up
Prostate	Musicco <i>et al.</i> , 2013	0.94 (0.56–1.46) [♦]	Age, sex, and calendar year of follow-up
Colorectal	Musicco <i>et al.</i> , 2013	0.43 (0.23–0.74) [♦]	Age, sex, and calendar year of follow-up
Non-melanoma skin cancer	Schmidt <i>et al.</i> , 2017	0.95 (0.92–0.98)	Age, sex, calendar period of cancer diagnosis, alcohol-related diagnosis, hospital-diagnosed obesity, hypertension, ischemic heart disease, congenitive heart failure, peripheral artery disease, chronic pulmonary disease, diabetes, cancer, and multiple sclerosis
Non-smoking related	Freedman <i>et al.</i> , 2016	0.86 (0.83–0.89)	Race, sex, and number of doctors' visits, birth year, and cancer registry area
Prostate	Chung <i>et al.</i> , 2016	1.71 (0.90–3.25)	Geographical location, monthly income, urbanization level, hypertension, diabetes, coronary heart disease, hyperlipidemia, and stroke
Non-melanoma skin cancer	White <i>et al.</i> , 2013	0.18 (0.024–1.34)	Sex, education, occupation, hypertension, diabetes, coronary heart disease, and ApoE4 status
Prostate	Frain <i>et al.</i> , 2017	1.08 (1.04–1.11)	Number of clinic visits in the year before baseline, follow-up time, cancer treatment, high cholesterol, hypertension, obesity, coronary arterial disease, diabetes, stroke, and race
Colorectal	Frain <i>et al.</i> , 2017	0.97 (0.90–1.05)	Number of clinic visits in the year before baseline, follow-up time, cancer treatment, high cholesterol, hypertension, obesity, coronary arterial disease, diabetes, stroke, and race
Melanoma	Frain <i>et al.</i> , 2017	1.11 (0.99–1.23)	Number of clinic visits in the year before baseline, follow-up time, cancer treatment, high cholesterol, hypertension, obesity, coronary arterial disease, diabetes, stroke, and race
Lymphoma	Frain <i>et al.</i> , 2017	0.86 (0.73–0.99)	Number of clinic visits in the year before baseline, follow-up time, cancer treatment, high cholesterol, hypertension, obesity, coronary arterial disease, diabetes, stroke, and race
Leukemia	Frain <i>et al.</i> , 2017	0.81 (0.68–0.97)	Number of clinic visits in the year before baseline, follow-up time, cancer treatment, high cholesterol, hypertension, obesity, coronary arterial disease, diabetes, stroke, and race
Myeloma	Frain <i>et al.</i> , 2017	0.80 (0.60–1.07)	Number of clinic visits in the year before baseline, follow-up time, cancer treatment, high cholesterol, hypertension, obesity, coronary arterial disease, diabetes, stroke, and race
Liver	Frain <i>et al.</i> , 2017	0.76 (0.46–1.23)	Number of clinic visits in the year before baseline, follow-up time, cancer treatment, high cholesterol, hypertension, obesity, coronary arterial disease, diabetes, stroke, and race
Any cancer			
Any	Driver <i>et al.</i> , 2012 [‡]	0.67 (0.47–0.97)	Age, sex, and smoking
Any	Musicco <i>et al.</i> , 2013	0.57 (0.49–0.67) [♦]	Age, sex, and calendar year of follow-up
Any	Realmuto <i>et al.</i> , 2012	0.60 (0.40–1.10) [‡]	Age, sex, education, and smoking
Any	Bowles <i>et al.</i> , 2017	0.95 (0.77–1.17)	Age at study entry, cohort, sex, education, diabetes, hypertension, heart disease, stroke, smoking status, low self-rated health, regular exercise, and body mass index
Any	Yarchoan <i>et al.</i> , 2017	0.73 (0.56–0.94) [‡]	Age, sex, education, race, and ApoE4 status
Any	Freedman <i>et al.</i> , 2016	0.87 (0.84–0.90)	Race, sex, and number of doctors' visits, birth year, and cancer registry area
Any	White <i>et al.</i> , 2013	0.95 (0.47–1.93)	Sex, education, occupation, hypertension, diabetes, coronary heart disease, and ApoE4 status
Any	Frain <i>et al.</i> , 2017	1.00 (0.97–1.03)	Number of clinic visits in the year before baseline, follow-up time, cancer treatment, high cholesterol, hypertension, obesity, coronary arterial disease, diabetes, stroke, and race
Any	Roe <i>et al.</i> , 2010	0.31 (0.12–0.86)	Sex, race, education, age, income, smoking, >130% overweight, kilocalories expended in physical activity, and Cardiovascular Health Study clinic effect

Table 1. Observational evidence for cancer and risk of subsequent Alzheimer's disease, by cancer type. Separate effect estimates are provided for smoking-related cancers, non-smoking related cancers, and all cancers together. Unless otherwise noted, the maximally adjusted effect estimate is displayed. *Smoking-related cancers include oral cavity and pharynx, lip, pancreas, lung and bronchus, larynx, cervix, kidney and renal pelvis, bladder, esophagus, and stomach. **Smoking-related cancers include oral, pharynx, larynx, esophagus, stomach, pancreas, lung, cervix, bladder, and kidney. ‡Maximally adjusted model was not chosen here, due to insufficient statistical power. ♦Relative risk †Odds ratio.

Cancer type	No. of SNPs	Odds ratio	Lower confidence limit	Upper confidence limit	P-value	BFDR _{50%}	BFDR _{10%}	BFDR _{1%}	BFDR _{0.1%}
Smoking-related cancers									
Renal cell carcinoma	12	0.97	0.89	1.05	0.434	68.22	95.08	99.53	99.95
Pancreatic cancer	13	0.97	0.91	1.04	0.429	71.63	95.79	99.60	99.96
Upper aerodigestive tract cancer	14	0.96	0.90	1.03	0.235	64.60	94.26	99.45	99.95
Urinary bladder cancer	11	0.94	0.83	1.06	0.325	58.76	92.77	99.30	99.93
Lung cancer	18	0.91	0.84	0.99	0.019	20.57	69.98	96.25	99.62
Smoking-related cancers (all)	68	0.95	0.92	0.98	0.003	7.11	40.78	88.33	98.71
Non-smoking related cancers									
Prostate cancer	36	0.98	0.95	1.01	0.234	77.05	96.80	99.70	99.97
Leukemia	38	0.98	0.96	0.995	0.012	34.13	82.35	98.08	99.81
Breast cancer	109	0.94	0.89	0.99	0.028	30.32	79.66	97.73	99.77
Melanoma	24	0.98	0.93	1.03	0.447	76.52	96.70	99.69	99.97
Lymphoma	20	1.01	0.95	1.07	0.764	79.89	97.28	99.75	99.98
Ovarian cancer	19	1.01	0.96	1.06	0.797	81.76	97.58	99.78	99.98
Non-smoking-related cancers (all)	246	0.98	0.97	0.995	0.009	34.62	82.65	98.13	99.81
Smoking/non-smoking related cancers									
All cancers	314	0.98	0.96	0.99	2.7E-4	2.19	16.74	68.86	95.71

Table 2. Results of conventional Mendelian randomization analysis for cancer and Alzheimer's disease. Genetically predicted lung cancer, leukemia, breast cancer, smoking-related cancers, non-smoking related cancers, and all cancers taken together were associated with significantly lower odds of Alzheimer's disease. The Bayes False Discovery Rate (BFDR) is also provided for each result as an approximation of the false positive rate. BFDR_{50%} = Bayes False Discovery Rate at prior of 50%, BFDR_{10%} = Bayes False Discovery Rate at prior of 10%, BFDR_{1%} = Bayes False Discovery Rate at prior of 1%, BFDR_{0.1%} = Bayes False Discovery Rate at prior of 0.1.

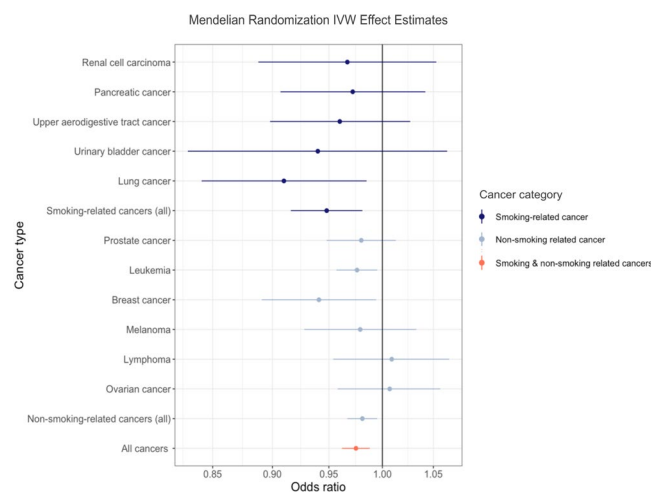


Figure 3. Odds ratio of Alzheimer's disease per genetically predicted increase in risk of cancer. Each circle represents the inverse-variance weighted Mendelian randomization estimate for the causal effect of the corresponding genetically predicted cancer on Alzheimer's disease. Dark blue represents smoking-related cancers, steel blue represents non-smoking related cancers, and orange represents all cancers in aggregate. Approximately 86% of points fell to the left of the null (OR = 1), while 14% fell to the right of the null. Genetically predicted lung cancer, leukemia, and breast cancer, as well as smoking-related cancers, non-smoking related cancers, and all cancers together were associated with significantly lower odds of Alzheimer's disease.

Chromosome (Chromosome of the SNP (Build 37, Assembly Hg19)), **Position** (position of the SNP (Build 37, Assembly Hg19)), **MarkerName** (SNP rsID or chromosome:position if rsID not available), **Effect allele** (reference allele (coded allele)), **Non Effect allele** (non-reference allele (non-coded allele)), **Beta-coefficient** (overall estimated effect size for the effect allele), **Standard error** (overall standard error for effect size estimate), **P-value** (meta-analysis p-value using regression coefficients (beta and standard error)).

The IGAP summary statistics corresponding to each cancer-related SNP are detailed in Supplementary Table 6.

Cancer type	Intercept	Standard error	P-value	Q	Q P-value
Lung cancer	0.024	0.014	0.118	16.79	0.209
Leukemia	-0.004	0.008	0.645	34.10	0.732
Breast cancer	-0.001	0.004	0.770	111.22	0.153
Smoking-related cancers	-0.003	0.008	0.695	84.73	0.006
Non-smoking related cancers	-0.002	0.002	0.254	265.01	0.038
All cancers	-0.003	0.002	0.167	353.53	0.002

Table 3. MR-Egger intercept test for unbalanced horizontal pleiotropy. In all cases, there is no evidence of unbalanced horizontal pleiotropy from the MR Egger Test. Cochran's Q statistic is significant for smoking-related, non-smoking related, and all cancers.

Two-sample mendelian randomization. *Inverse-variance weighted effect estimates and false discovery rate.* Genetically predicted smoking-related cancers were associated with 5.2% lower odds of AD (OR 0.95, 95% CI 0.92 to 0.98, $p = 0.0026$; $\text{BFDR}_{50\%} = 7.11\%$, $\text{BFDR}_{10\%} = 40.78\%$, $\text{BFDR}_{1\%} = 88.34\%$, $\text{BFDR}_{0.1\%} = 98.71\%$) per 1-unit higher log odds of cancer (Table 2 and Fig. 3). Within smoking-related cancers, only genetically predicted lung cancer, with a 9.0% reduction, was associated with AD (OR 0.91, 95% CI 0.84 to 0.99, $p = 0.019$; $\text{BFDR}_{50\%} = 20.57\%$, $\text{BFDR}_{10\%} = 69.98\%$, $\text{BFDR}_{1\%} = 96.25\%$, $\text{BFDR}_{0.1\%} = 99.62\%$) per 1-unit higher log odds of cancer (Table 2 and Fig. 3). As only one susceptibility locus was identified for cervical cancer, separate Mendelian randomization analyses were not conducted for cervical cancer; however, its corresponding SNP was used in the pooled analysis of all smoking-related cancers.

Genetically predicted non-smoking related cancers were associated with 1.9% lower odds of AD (OR 0.98, 95% CI 0.97 to 0.995, $p = 0.0091$; $\text{BFDR}_{50\%} = 34.62\%$, $\text{BFDR}_{10\%} = 82.65\%$, $\text{BFDR}_{1\%} = 98.13\%$, $\text{BFDR}_{0.1\%} = 99.81\%$) per 1-unit higher log odds of cancer (Table 2 and Fig. 3). Within non-smoking related cancers, genetically predicted prostate cancer, melanoma, lymphoma, and ovarian cancer were not associated with differential odds of developing AD. However, genetically predicted leukemia was associated with 2.4% reduced odds of AD (OR 0.98, 95% CI 0.96 to 0.995, $p = 0.012$; $\text{BFDR}_{50\%} = 34.13\%$, $\text{BFDR}_{10\%} = 82.34\%$, $\text{BFDR}_{1\%} = 98.09\%$, $\text{BFDR}_{0.1\%} = 99.81\%$) per 1-unit higher log odds of cancer, and breast cancer was associated with 5.9% lower odds of AD (OR 0.94, 95% CI 0.89 to 0.99, $p = 0.028$; $\text{BFDR}_{50\%} = 30.32\%$, $\text{BFDR}_{10\%} = 79.66\%$, $\text{BFDR}_{1\%} = 97.73\%$, $\text{BFDR}_{0.1\%} = 99.77\%$) per 1-unit higher log odds of cancer (Table 2 and Fig. 3).

When genetic predictors of cancers at all sites under consideration were pooled, cancer was associated with a 2.5% reduction in odds of AD (OR 0.98, 95% CI 0.96 to 0.99, $p = 0.00027$; $\text{BFDR}_{50\%} = 2.18\%$, $\text{BFDR}_{10\%} = 16.74\%$, $\text{BFDR}_{1\%} = 68.86\%$, $\text{BFDR}_{0.1\%} = 95.71\%$) per 1-unit higher log odds of cancer (Table 2 and Fig. 3).

Sensitivity analyses. MR-Egger intercept test: Pleiotropy was assessed based on the intercept of MR-Egger analysis and Cochran's Q. In all cases, there was no evidence to reject the null hypothesis of no unmeasured directional pleiotropy of the genetic variants in MR-Egger analyses (i.e. MR-Egger intercept of 0) (Table 3). There was evidence of heterogeneity from Cochran's Q statistic for smoking-related, non-smoking related, and all cancers (Table 3).

Leave-one-out analysis: To further examine the stability of the Mendelian randomization effect estimates and identify any important outliers, each SNP was sequentially removed from the analysis, conducting Y analyses with $Y-1$ datapoints. The precision and direction of the association between cancer and risk of AD remained largely unaltered with this approach. Only four SNPs had a marginal influence on the overall estimate (namely, for the association between lung cancer and AD and breast cancer and AD). Plots of leave-one-out analysis showing the influence of individual SNPs on the overall effect estimate for each cancer type and risk of AD are provided in Appendix 2.

Funnel plots of inverse standard error: Any heterogeneity of effect estimates, and thereby the likelihood of directional pleiotropy of SNPs, was assessed on the basis of funnel plots depicting the inverse standard error of the causal estimate for each genetic variant. Asymmetry about the vertical line is indicative of heterogeneity, possibly due to pleiotropy. Figure 4 depicts the funnel plots for lung cancer, smoking-related cancers, leukemia, breast cancer, non-smoking related cancers, and all cancers.

Discussion

Our results are consistent with a protective effect of some cancer-related gene variants against Alzheimer's disease, and support genetically mediated mechanisms for the inverse epidemiological association. Specifically, genes associated with increased risk of cancer at any site were associated with a reduced risk of Alzheimer's disease, and across individual cancer types, lung cancer, breast cancer, and leukemia were found to be associated with a statistically significant lower risk of AD. Finally, genetically predicted smoking-related cancers were more strongly associated with AD than non-smoking related cancers. This Mendelian randomization approach provides critical information on causality, beyond the former epidemiological observations, with important implications for public health and the prevention of both cancer and AD.

Taken together, the results of our Mendelian randomization complement and extend existing observational evidence of an inverse relationship between cancer and AD. Two studies synthesized observational data covering multiple cancer types to compare the relationship between smoking-related cancers and AD with the relationship between non-smoking related cancers and AD^{6,11}. In line with their findings, our study demonstrates an inverse causal relationship between genetically predicted cancers and risk of AD and, furthermore, supports a

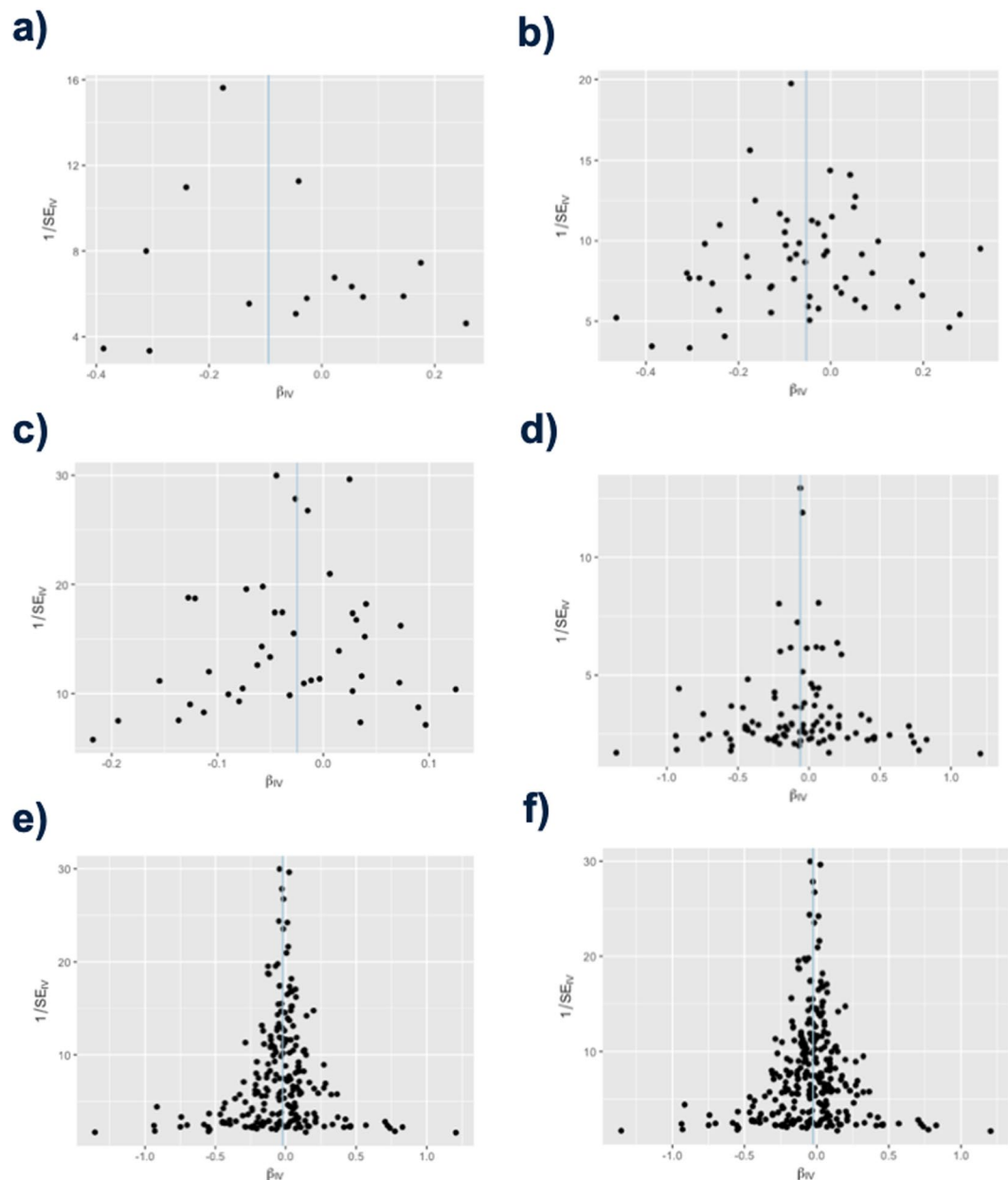


Figure 4. Funnel plots depicting the relationship between the causal effect of cancer on Alzheimer's disease estimated by each genetic variant against the inverse standard error of the causal estimate. The x-axis represents the effect estimate (beta-coefficient) for risk of Alzheimer's disease for each SNP. The y-axis represents the standard error of the estimated effect. The blue vertical line represents the inverse-variance weighted Mendelian randomization estimate using all SNPs. In all cases, the estimated effects scatter roughly symmetrically about the overall Mendelian randomization estimate, indicating precision in the estimated effects as well as absence of pleiotropic effects. **(a)** lung cancer **(b)** smoking-related cancers **(c)** leukemia **(d)** breast cancer **(e)** non-smoking related cancers **(f)** all cancers.

stronger relationship between genetically predicted smoking-related cancers and AD than genetically predicted non-smoking related cancers and AD.

Strengths of this study include the use of data from large GWAS of cancer, as well as the MR design, which reduces confounding and bias from reverse-causality. Furthermore, the validity of MR relies on the assumptions that the instrumental variables are strongly associated with the risk factor of interest, are not associated with any confounders, and do not influence the outcome through alternative causal pathways (Fig. 5)³⁵. These factors were addressed in the study design³⁶. While relying on published summary statistics for our analyses means that instrumental variable strengths cannot be formally evaluated through the calculation of an F-statistic³⁷, any instrumental bias would, in this case, be in the direction of the null, given that analyses were undertaken in separate settings for the exposure and outcome³⁸. Therefore, bias from weak instruments cannot explain the observed relationship between cancer and AD. Furthermore, while the wide range of risk factors for AD and the complex,

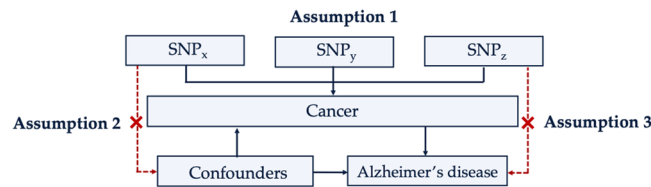


Figure 5. Core assumptions of Mendelian randomization. The validity of Mendelian randomization relies on three core assumptions: (1) the genetic variants are associated with the exposure (cancer), (2) the genetic variants are not associated with any confounders, and (3) the genetic variants influence the risk of the outcome (Alzheimer's disease) through the exposure (cancer), and not any other pathways. SNP_x, SNP_y, SNP_z = single nucleotide polymorphisms. Adapted from Larsson *et al.*, 2017.

interlaced nature of physiological signaling complicate this assumption, results of the MR-egger intercept test, as discussed above, demonstrate that pleiotropy is unlikely to explain the observed association between cancer and AD. While there was evidence for heterogeneity of effect from Cochran's Q statistics for some cancers, there was little to suggest that specific SNPs contributed to this heterogeneity based on leave-one-out analyses. Moreover, measurement error in the exposure can lead to heterogeneity when no pleiotropy is present³⁹.

Nonetheless, our results must be considered in the context of several important limitations. An important potential limitation is the variability in the diagnostic criteria used to identify AD in IGAP studies and the inevitable possibility of misclassification bias. Another potential source of bias in Mendelian randomization analysis is population stratification. However, the IGAP dataset is restricted to individuals of European ancestry, thus reducing this possibility. Finally, cancer is a binary causal exposure that can be considered as a continuous variable representing cancer risk. It has been shown that a binary exposure that is a dichotomization of a continuous risk factor can lead to violation of the exclusion restriction assumption and limit the inferences drawn from an MR study⁴⁰. However, this only affects the causal effect estimate and, provided the instrumental variable assumptions are satisfied, the use of MR to test the causal null hypothesis remains valid.

We present largely concordant results from Bayesian and frequentist statistical analyses, while recognizing the very different nature of the statistical tests. The Bayesian analysis quantifies the relative probability of alternative hypotheses (including a null hypothesis) given the observed data distributions. In view of the consistency of existing observational evidence for smoking-related cancers, a BFDR prior of 50% was used, while a more stringent prior of 10% was used to evaluate non-smoking related cancers. Under these assumptions, a caveat is necessary for the interpretation of the apparent associations for breast cancer, leukemia, and non-smoking related cancers (noting the BFDR = 30% for breast cancer, 34% for leukemia, and 35% for non-smoking related cancers). However, when all cancers are considered in aggregate, or just smoking-associated cancers, the Mendelian randomization approach indicates that genes associated with cancer are associated with lower risk of AD. Twenty-eight genetic variants linked to cancer were associated with lower odds of AD, 13 of which are located either on or near genes that have been previously linked with AD (Supplementary Table 6). The remaining 15 SNPs merit further study, as they may contribute in developing novel diagnostic and prognostic tools for AD, as well as effective treatments strategies for the condition.

Expansion of this study and replication in other ethnic cohorts will be necessary to determine the external validity of our findings. Future studies should additionally seek to investigate the relationship between cancer and other forms of neurodegeneration, broadly. Ultimately, these results may provide grounds for cautious optimism about the prospects for drug repurposing from cancer to AD—which may help to shorten the timelines for dementia drug discovery—and emphasize that genetic studies can help to deconvolute the complex interrelation between these two disorders.

References

- Ma, L. L. *et al.* Association between cancer and Alzheimer's disease: systematic review and meta-analysis. *Journal of Alzheimer's disease: JAD* **42**, 565–573, <https://doi.org/10.3233/jad-140168> (2014).
- Papageorgakopoulos, T. N., Moraitou, D., Papanikolaou, M. & Tsolaki, M. The association between Alzheimer's disease and cancer: Systematic review - Meta-analysis. *Hellenic journal of nuclear medicine* **20**(Suppl), 45–57 (2017).
- Zhang, Q. *et al.* Inverse relationship between cancer and Alzheimer's disease: a systemic review meta-analysis. *Neurological sciences: official journal of the Italian Neurological Society and of the Italian Society of Clinical Neurophysiology* **36**, 1987–1994, <https://doi.org/10.1007/s10072-015-2282-2> (2015).
- Houck, A. L., Seddighi, S. & Driver, J. A. At the Crossroads Between Neurodegeneration and Cancer: A Review of Overlapping Biology and Its Implications. *Current aging science*, <https://doi.org/10.2174/1874609811666180223154436> (2018).
- Roe, C. M. *et al.* Cancer linked to Alzheimer disease but not vascular dementia. *Neurology* **74**, 106–112 (2010).
- Freedman, D. M. *et al.* Associations between cancer and Alzheimer's disease in a U.S. Medicare population. *Cancer Medicine* **5**, 2965–2976, <https://doi.org/10.1002/cam4.850> (2016).
- White, R. S., Lipton, R. B., Hall, C. B. & Steiner, J. R. Nonmelanoma skin cancer is associated with reduced Alzheimer disease risk. *Neurology* **80**, 1966–1972, <https://doi.org/10.1212/WNL.0b013e3182941990> (2013).
- Realmutto, S. *et al.* Tumor diagnosis preceding Alzheimer's disease onset: is there a link between cancer and Alzheimer's disease? *Journal of Alzheimer's disease: JAD* **31**, 177–182, <https://doi.org/10.3233/jad-2012-120184> (2012).
- Frain, L. *et al.* Association of cancer and Alzheimer's disease risk in a national cohort of veterans. *Alzheimer's & Dementia* **13**, 1364–1370 (2017).
- Hanson, H. A., Horn, K. P., Rasmussen, K. M., Hoffman, J. M. & Smith, K. R. Is Cancer Protective for Subsequent Alzheimer's Disease Risk? Evidence From the Utah Population. *Database. The journals of gerontology. Series B, Psychological sciences and social sciences* **72**, 1032–1043, <https://doi.org/10.1093/geronb/gbw040> (2017).

11. Driver, J. A. *et al.* Inverse association between cancer and Alzheimer's disease: results from the Framingham Heart Study. *BMJ (Clinical research ed.)* **344** (2012).
12. Roe, C. M., Behrens, M. I., Xiong, C., Miller, J. P. & Morris, J. C. Alzheimer disease and cancer. *Neurology* **64**, 895–898, <https://doi.org/10.1212/01.Wnl.0000152889.94785.51> (2005).
13. Musicco, M. *et al.* Inverse occurrence of cancer and Alzheimer disease: a population-based incidence study. *Neurology* **81**, 322–328, <https://doi.org/10.1212/WNL.0b013e31829c5ec1> (2013).
14. Burgess, S., Timpson, N. J., Ebrahim, S. & Davey Smith, G. Mendelian randomization: where are we now and where are we going? *International Journal of Epidemiology* **44**, 379–388, <https://doi.org/10.1093/ije/dyv108> (2015).
15. Qi, L. Mendelian randomization in nutritional epidemiology. *Nutrition reviews* **67**, 439–450, <https://doi.org/10.1111/j.1753-4887.2009.00218.x> (2009).
16. MacArthur, J. *et al.* The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic acids research* **45**, D896–D901 (2016).
17. Lambert, J. *et al.* European Alzheimer's Disease Initiative Investigators. Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. *Nat Genet* **41**, 1094–1099 (2009).
18. Lambert, J.-C. *et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nature genetics* **45**, 1452 (2013).
19. Howie, B. N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS genetics* **5**, e1000529 (2009).
20. Li, Y., Willer, C. J., Ding, J., Scheet, P. & Abecasis, G. R. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genetic epidemiology* **34**, 816–834 (2010).
21. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nature genetics* **39**, 906 (2007).
22. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics* **81**, 559–575 (2007).
23. Aulchenko, Y. S., Struchalin, M. V. & van Duijn, C. M. ProbABEL package for genome-wide association analysis of imputed data. *BMC bioinformatics* **11**, 134 (2010).
24. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
25. Mägi, R. & Morris, A. P. GWAMA: software for genome-wide association meta-analysis. *BMC bioinformatics* **11**, 288 (2010).
26. Hartwig, F. P., Davies, N. M., Hemani, G. & Davey Smith, G. Two-sample Mendelian randomization: avoiding the downsides of a powerful, widely applicable but potentially fallible technique. *Int J Epidemiol.* **45**(6), 1717–1726, <https://doi.org/10.1093/ije/dyx028> (2016).
27. Burgess, S. & Thompson, S. G. Interpreting findings from Mendelian randomization using the MR-Egger method. *European Journal of Epidemiology* **32**, 377–389, <https://doi.org/10.1007/s10654-017-0255-x> (2017).
28. Hemani, G. *et al.* The MR-Base platform supports systematic causal inference across the human phenotype. *Elife* **7**, e34408 (2018).
29. Bowden, J. *et al.* Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I² statistic. *International journal of epidemiology* **45**, 1961–1974, <https://doi.org/10.1093/ije/dyw220> (2016).
30. Schmidt, S. A. J., Ording, A. G., Horváth-Puhó, E., Sørensen, H. T. & Henderson, V. W. Non-melanoma skin cancer and risk of Alzheimer's disease and all-cause dementia. *PLoS one* **12**, e0171527, <https://doi.org/10.1371/journal.pone.0171527> (2017).
31. Chung, M. S. & Lee, S. H. Current status of active surveillance in prostate cancer. *Investigative and Clinical Urology* **57**, 14–20, <https://doi.org/10.4111/icu.2016.57.1.14> (2016).
32. Bowles, E. J. A. *et al.* Risk of Alzheimer's disease or dementia following a cancer diagnosis. *PLoS one* **12**, e0179857, <https://doi.org/10.1371/journal.pone.0179857> (2017).
33. Yarchoan, M. *et al.* Association of Cancer History with Alzheimer's Disease Dementia and Neuropathology. *Journal of Alzheimer's disease: JAD* **56**, 699–706, <https://doi.org/10.3233/JAD-160977> (2017).
34. Ma, C., Liu, Y., Neumann, S. & Gao, X. Nicotine from cigarette smoking and diet and Parkinson disease: a review. *Translational neurodegeneration* **6**, 18 (2017).
35. Larsson, S. C. *et al.* Modifiable pathways in Alzheimer's disease: Mendelian randomisation analysis. *BMJ (Clinical research ed.)* **359**, j5375 (2017).
36. Davey Smith, G. & Hemani, G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Human molecular genetics* **23**, R89–R98, <https://doi.org/10.1093/hmg/ddu328> (2014).
37. Burgess, S. & Thompson, S. G. Avoiding bias from weak instruments in Mendelian randomization studies. *International journal of epidemiology* **40**, 755–764, <https://doi.org/10.1093/ije/dyr036> (2011).
38. Pierce, B. L., Ahsan, H. & VanderWeele, T. J. Power and instrument strength requirements for Mendelian randomization studies using multiple genetic variants. *International journal of epidemiology* **40**, 740–752, <https://doi.org/10.1093/ije/dyq151> (2011).
39. Hemani, G., Bowden, J. & Davey Smith, G. Evaluating the potential role of pleiotropy in Mendelian randomization studies. *Human molecular genetics* **27**, R195–R208 (2018).
40. Burgess, S. & Labrecque, J. A. Mendelian randomization with a binary exposure variable: interpretation and presentation of causal estimates. *European Journal of Epidemiology* **33**, 947–952, <https://doi.org/10.1007/s10654-018-0424-6> (2018).

Acknowledgements

We would like to thank those involved in the International Genomics of Alzheimer's Project (IGAP) for the summary statistics used within this paper. IGAP investigators contributed to the design and implementation of their study and/or provided data, but did not participate in the analysis or writing of the present study. This project was supported by a Gates Cambridge Trust grant to SS (OPP1144). JBR is supported by the Wellcome Trust (103838). PDPP has infrastructural support from Cambridge University and Cancer Research UK.

Author Contributions

S.S. analyzed the data, drew the figures, and wrote the first draft of the manuscript. S.S. and P.D.P. were involved in the conception and design of the study. All authors contributed to the interpretation of the results and revision of the manuscript and approved the final version. For the International Genomics of Alzheimer's Project, the investigators contributed to the design and implementation of their study and/or provided data, but did not participate in the analysis or writing of the present study.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-019-49795-6>.

Competing Interests: The authors declare no competing interests.

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019