

Association between lifetime smoking and cutaneous squamous cell carcinoma: A 2-sample Mendelian randomization study



Truelian Lee, BA,^a Christopher D. George, MD,^b Chen Jiang, PhD,^c Maryam M. Asgari, MD, MPH,^d Tamar Nijsten, MD, PhD,^b Luba M. Pardo, MD, PhD,^b and Hélène Choquet, PhD^c

Background/Purpose: Cutaneous squamous cell carcinoma (cSCC) is one of the most common malignancies worldwide. While several environmental risk factors for cSCC are well established, there is conflicting evidence on cigarette smoking (and its potential causal effect) and cSCC risk. Furthermore, it is unclear if these potential associations represent causal, modifiable risk factors for cSCC development. This study aims to assess the nature of the associations between cigarette smoking traits (smoking initiation, amount smoked, and lifetime smoking exposure) and cSCC risk using two-sample Mendelian randomization analyses.

Methods: Genetic instruments, based on common genetic variants associated with cigarette smoking traits ($P < 5 \times 10^{-8}$), were derived from published genome-wide association studies (GWASs). For cSCC, we used GWAS summary statistics from the Kaiser Permanente GERA cohort (7701 cSCC cases and 60,167 controls; all non-Hispanic Whites).

Results: We found modest evidence that genetically determined lifetime smoking was associated with cSCC (inverse-variance weighted method: OR[95% CI] = 1.47[1.09-1.98]; $P = .012$), suggesting it may be a causal risk factor for cSCC. We did not detect any evidence of association between genetically determined smoking initiation or amount smoked and cSCC risk.

Conclusion: Study findings highlight the importance of smoking prevention and may support risk-stratified cSCC screening strategies based on carcinogen exposure and other genetic and clinical information. (JAAD Int 2024;14:69-76.)

Key words: cigarette smoking; cutaneous squamous cell carcinoma; genetic epidemiology; Mendelian randomization.

INTRODUCTION

Cutaneous squamous cell carcinoma (cSCC) is one of the most common skin cancers worldwide, with an increasing prevalence in recent years.¹ cSCCs frequently occur in sun-exposed areas such as the

head and neck.² According to the Global Burden of Disease (GBD) study, the cutaneous malignancy with the largest worldwide increase in prevalence between 1990 and 2017 was cSCC with a 310% increase.¹ In a more recent GBD analysis, there

From the Harvard Medical School, Boston, Massachusetts^a; Department of Dermatology, Erasmus MC Cancer Institute, University Medical Center Rotterdam, Rotterdam, Netherlands^b; Division of Research, Kaiser Permanente Northern California, Oakland, California^c; and Department of Dermatology, University of Colorado Anschutz Medical Campus, Aurora, Colorado.^d

Drs Lee and George contributed equally to this article.

Funding sources: H.C. and M.M.A. are supported by the National Cancer Institute (NCI) grant R01CA241623. The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

IRB approval status: The Institutional Review Board of the Kaiser Foundation Research Institute approved all study procedures, and written informed consent was obtained from all participants.

Accepted for publication November 22, 2023.

Correspondence to: Hélène Choquet, PhD, Division of Research, Kaiser Permanente Northern California, 2000 Broadway, Oakland, CA 94612. E-mail: Helene.Choquet@kp.org.

2666-3287

© 2023 by the American Academy of Dermatology, Inc. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.1016/j.jdin.2023.11.005>

was an increase in the global all-age disability-adjusted life years for cSCC in both sexes combined, compared to that in 1990 and 2010.³

Besides well-established risk factors for cSCC, including genetic risk factors, such as those affecting pigmentation traits, and environmental factors (e.g., UV radiation exposure and immunosuppression),⁴ cigarette smoking is a potential risk factor for cSCC, as it has been associated with a wide variety of neoplasms.⁵⁻⁹ Despite prior observational studies demonstrating a link between smoking and cSCC, the causal association between the 2 remains unclear.¹⁰⁻¹² Additional limitations to prior analyses include selection bias (nonparticipation could be related to smoking amount/status or people who smoke may have higher cumulative UV exposure levels) and difficulty in accurately assessing the amount and duration of smoking (certain studies did not distinguish between cigar and cigarettes, or former, never, and current smokers).¹⁰⁻¹²

Compared to observational studies, Mendelian randomization (MR) studies help to assess the causal relation between an exposure and an outcome. In MR, genetic variants are used as instrumental variables (IVs), with the assumption that humans are all randomly assigned genetic variants from their parents.¹³ This is significant because genetic variants, if associated with the exposure and not directly with the outcome and not associated with confounders, can serve as a reliable IV, hence enabling causal inference. Because of the randomization in genetic variants in a population, MR can reduce the effect of confounding, reverse causation, and various biases seen in observational studies.¹³ Although the application of MR rests upon more extensive assumptions than a randomized controlled trial (RCT), MR can be used in cases where randomized controlled trials (RCTs) are not feasible or ethical.^{13,14}

The present study aims to elucidate the causal association of cigarette smoking traits (ie, smoking initiation, amount smoked, and lifetime smoking exposure) with cSCC risk, and circumvent prior limitations, by using a two-sample MR approach. We used 3 separate cigarette smoking traits as the exposures, as those represent different stages of cigarette use: initiation for “smoking initiation,” heaviness for “amount smoked,” and initiation,

duration, heaviness, and cessation for “lifetime smoking.” We compare genetic effect estimates for those 3 cigarette smoking traits (exposures) and cSCC risk (outcome) obtained through GWAS summary statistics, especially, from our previous GWAS of cSCC conducted in the Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort.¹⁵ Through our study, we aim to highlight the potential value of integrating the desire to smoke (determined genetically) in cSCC risk assessment, thereby informing targeted prevention and screening strategies for populations at higher risk of developing the disease.

METHODS

Study design

Two-sample MR analyses were conducted to investigate separately the association of

genetically determined smoking initiation, genetically determined amount smoked, and genetically determined lifetime smoking exposure with the risk of cSCC. For each of the 3 exposures, we used the lead single-nucleotide polymorphisms (SNPs) previously reported as genome-wide significant ($P < 5.0 \times 10^{-8}$) as a set of genetic instruments. Genetic instruments were then clumped using a window of 10 Mb and maximal linkage disequilibrium of $r^2 = 0.001$ between instruments to ensure that genetic variants were independent. The different data sets used for this MR study are summarized in Supplementary Table I, available via Mendeley at <https://doi.org/10.17632/cnkvc82bgr.1>.

GWAS summary statistics for cSCC

Genetic association data for cSCC risk (outcome) were retrieved from our previous GWAS study conducted in the GERA cohort.¹⁵ The GERA cohort consists of 110,266 adult members of the Kaiser Permanente Medical Care Plan, Northern California Region (KPNC), an integrated healthcare delivery system, that includes ongoing longitudinal electronic health records (EHRs).^{16,17} The Institutional Review Board (IRB) of the Kaiser Foundation Research Institute approved all study procedures. Written informed consent was obtained from all participants. In the current study, we retrieved genetic association data from the GWAS of cSCC conducted in 7701 SCC cases and 60,167 controls; all GERA participants of European ancestry.¹⁵

CAPSULE SUMMARY

- Considering conflicting evidence regarding the association between cigarette smoking and cSCC, we investigated if there was genetic evidence of a potential causal relationship between cigarette smoking traits and cSCC risk.
- Our findings provide genetic evidence that increased lifetime smoking may be a causal risk factor for cSCC.

Abbreviations used:

cSCC:	cutaneous squamous cell carcinoma
GERA:	Genetic Epidemiology Research on Adult Health and Aging
GSCAN:	GWAS and Sequencing Consortium of Alcohol and Nicotine use
GWAS:	genome-wide association study
MR:	Mendelian randomization
SNP:	single nucleotide polymorphism
UKB:	UK Biobank

Genetic instruments for cigarette smoking traits

Genetic variants as IVs for cigarette smoking initiation (ever having smoked regularly versus never) and amount smoked (number of cigarettes per day) were extracted from the most recent GWAS and Sequencing Consortium of Alcohol and Nicotine use (GSCAN) study.¹⁸ GWAS summary statistics for smoking initiation and amount smoked analyses,¹⁸ included 805,431 and 326,497 individuals of European ancestry, respectively, from 42 cohorts (Supplementary Table II, available via Mendeley at <https://doi.org/10.17632/cnkvc82bgr.1>). Those GWAS summary statistics were publicly accessible at <https://conservancy.umn.edu/handle/11299/241912>. After clumping, a total of 236 genetic instruments for smoking initiation and 45 for cigarettes per day were used for the MR analyses (Supplementary Tables III-IV, available via Mendeley at xxx). Thus, by using these genetic variants, we adhered to the key assumption that the IVs are robustly associated with the exposure, as those genetic variants were previously reported as genome-wide significant in the large GSCAN study,¹⁸ making those strong IVs for cigarette smoking initiation and cigarettes per day. Moreover, the genetic scores of the 2 smoking-related traits have been previously reported to be associated with self-reported smoking behaviors in the GERA cohort.¹⁹

We also used genetic variants as instrumental variables for lifetime smoking (represented by an index which captures smoking status, duration, heaviness, and cessation) from a GWAS conducted in 462,690 UK Biobank (UKB) participants of European ancestry (54% female; mean age (SD) = 56.7 (8.0) years; 54% had never smoked).²⁰ The UKB is a longitudinal study following the health of approximately 500,000 participants aged 40-69 years, recruited from across the United Kingdom between 2006 and 2010.²¹ As previously described,²⁰ smoking measures available in UKB (ie, smoking status, age at initiation and at cessation, and number of cigarettes smoked per day) were self-reported and collected at initial assessment; smoking measures were then combined into a lifetime smoking index.

GWAS summary statistics for this study²⁰ were publicly accessible via GWAS Catalog under study accession identifier GCST009096. After clumping, a total of 121 genetic instruments for lifetime smoking were used for the MR analyses (Supplementary Table V, available via Mendeley at <https://doi.org/10.17632/cnkvc82bgr.1>).

Genome-wide genetic correlation analyses

We assessed genetic correlations (r_g) between the 3 cigarette smoking traits using cross-trait linkage disequilibrium score regression²² and using the above-mentioned GWAS summary statistics for smoking initiation and cigarettes per day¹⁸ and for lifetime smoking.²⁰

Two-sample MR analyses

All analyses were conducted in the R software (V.4.0.1) using the “TwoSampleMR” package.²³ This package makes causal inference about an exposure on an outcome using GWAS summary statistics, generates LD pruning of exposure SNPs, and harmonizes exposure and outcome data sets. We used the inverse-variance weighted (IVW) method as our primary source of MR estimates. This IVW method essentially translates to a weighted regression of SNP outcome effects on SNP-exposure effects where the intercept is constrained to zero. Moreover, we reported the estimations from MR weighted median, weighted mode, and MR-Egger. Furthermore, leave-one-SNP-out analyses were conducted (Supplementary Tables VI-VIII, available via Mendeley at <https://doi.org/10.17632/cnkvc82bgr.1>).

Sensitivity analyses

The potential effect of pleiotropy was evaluated by the regression intercept from the MR-Egger method²⁴ and Cochran Q tests were used to evaluate the presence of global heterogeneity amongst the effects of the genetic instruments²⁵ (Supplementary Table IX, available via Mendeley at <https://doi.org/10.17632/cnkvc82bgr.1>). The MR-PRESSO^{25,26} method was also used to provide an MR estimate which is robust against the presence of heterogeneity among SNP effects and to re-assess the MR estimate after excluding outlier SNPs.

RESULTS

Lifetime smoking exposure shares genetic determinants with smoking initiation and cigarettes per day

To quantify genetic overlap between the 3 exposures, genome-wide genetic correlation analyses were performed using cross-trait linkage

Table I. MR results of the associations of genetically predicted cigarette smoking traits with cSCC risk

Exposure (source)	Outcome	N genetic instruments	MR method	OR (95%CI)	P-value	Detected outlier SNP via MR-PRESSO
Cigarette smoking initiation (GSCAN)	cSCC (GERA)	236	IVW	0.99 (0.79-1.25)	.94	
		235	MR-PRESSO model	0.94 (0.86-1.26)	.69	rs7195043
Cigarettes per day (GSCAN)	cSCC (GERA)	45	IVW	0.91 (0.68-1.21)	.50	
		44	MR-PRESSO model	0.96 (0.75-1.23)	.76	rs2016968
Lifetime smoking (UKB)	cSCC (GERA)	121	IVW	1.47 (1.09-1.98)	.012	
		121	MR-PRESSO model	1.46 (1.09-1.97)	.013	NA

P-values reported in this table represent the statistical significance of the association between the exposure of interest (genetically determined smoking trait) and the outcome (cSCC risk). A threshold of $P < .05$ was used to determine significance and bold P-values are considered significant.

CI, Confidence interval; cSCC, cutaneous squamous cell carcinoma; GERA, Genetic Epidemiology Research on Adult Health and Aging; GSCAN, GWAS and Sequencing Consortium of Alcohol and Nicotine use; IVW, inverse-variance weighted model; MR, Mendelian randomization; OR, odds ratio; SNP, single-nucleotide polymorphism; UKB, UK Biobank.

disequilibrium score regression.²² Lifetime smoking exposure was genetically associated with smoking initiation (r_g , 0.87; SE, 0.01; $P = 1.0 \times 10^{-300}$) and cigarettes per day (r_g , 0.52; SE, 0.02; $P = 1.28 \times 10^{-165}$). Consistently, we also found evidence of genetic association between cigarettes per day and cigarette smoking initiation (r_g , 0.26; SE, 0.03; $P = 5.57 \times 10^{-19}$), as previously reported in the GSCAN study.¹⁸

Mendelian randomization analyses

We conducted two-sample MR analyses to investigate whether cigarette smoking traits causally influenced cSCC risk. No significant association was found between smoking initiation or the number of cigarettes smoked per day and cSCC risk (Table I and Supplementary Figs 1 and 2, available via Mendeley at <https://doi.org/10.17632/cnkvc82bgr.1>). In contrast, we found evidence for a causal effect of lifetime smoking exposure on cSCC risk, as an increase in lifetime smoking exposure was associated with an increased risk of cSCC (IVW model: odds ratio [OR] per one-unit increase = 1.47; 95% CI, 1.09-1.98; $P = .012$) (Table I and Fig 1).

Sensitivity analyses

No evidence of directional or horizontal pleiotropy was observed for all the analyses, as indicated by MR-Egger intercept P-values greater than 0.05. Furthermore, using the Cochran Q statistic, no significant heterogeneity was observed among the effects of the genetic instruments for lifetime smoking exposure ($Q = 117.58$, $P = .55$) (Supplementary Table XI, available via Mendeley at <https://doi.org/10.17632/cnkvc82bgr.1>). Finally, the MR-PRESSO test detected no outliers for lifetime smoking exposure (Table I).

DISCUSSION

Using MR on GWAS summary statistics from the Kaiser Permanente GERA cohort (7701 cSCC cases and 60,167 controls, all non-Hispanic whites), we found modest evidence that genetically determined lifetime smoking was significantly associated with cSCC, though there was no evidence of an association between genetically determined smoking initiation and amount smoked (cigarettes per day).

Our results support the findings of previous observational studies, which reported heterogeneous relationships of smoking traits with cSCC risk, depending on the trait tested and study population.^{10-12,27} Although our study used GWAS summary statistics from the largest study of smoking traits published to date (GSCAN consortium),¹⁸ which provided powerful genetic instruments for MR analyses, we did not observe evidence of causal relationship between smoking initiation or cigarettes per day and cSCC risk. In contrast, we found that genetically determined lifetime smoking was significantly associated with an increased risk of cSCC.

Our MR analyses were performed using valid IVs for causal inference under the 3 assumptions required for MR studies.^{28,29} The first assumption (ie, the IV should be truly associated with the exposure) was satisfied by the use of genetic variants previously reported as genome-wide significant in large studies,^{18,20} making those strong genetic instruments for cigarette smoking traits. The second assumption (ie, the IV should not be influenced by any confounders of the exposure-outcome association) was partially satisfied by the fact that we found similar results using the MR-PRESSO method, which is a robust method for sensitivity analysis. The third assumption (ie, the IV should only be related to the outcome of interest through the exposure under study) was satisfied because no evidence of

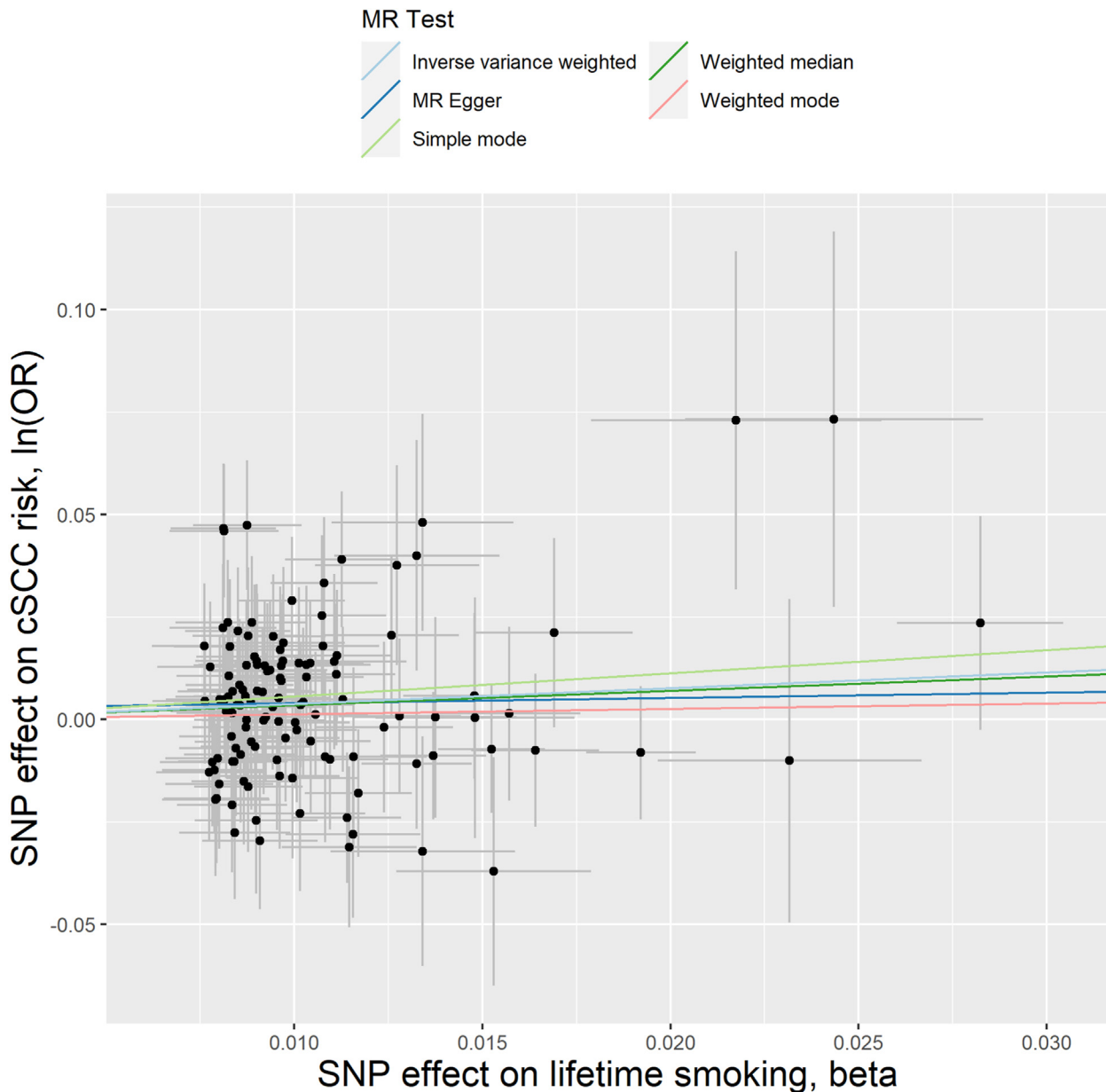


Fig 1. Association of lifetime smoking variants with the risk of cSCC. The x-axis shows 121 genetic instruments for lifetime smoking and their effect size estimates (ORs) with lifetime smoking. The y-axis shows the association of the same variants with cSCC risk. The Mendelian randomization (MR) inverse-weighted (IVW) regression line is plotted, along with MR Egger, simple mode, and weighted median model. cSCC, Cutaneous squamous cell carcinoma; OR, odds ratio; SNP, single-nucleotide polymorphism.

horizontal pleiotropy was detected. Altogether, our findings were unlikely to be affected by the violation of MR assumptions. Nevertheless, recent MR studies^{30,31} suggested that existing methods for detecting and accounting for horizontal pleiotropy are ineffective under some plausible conditions. Furthermore, genetic instruments for complex behavioral factors such as smoking traits seem to demonstrate horizontal pleiotropy.³² Future MR

investigations using genetic instruments for smoking traits may include negative control outcomes (ie, an outcome for which it is believed that the exposure cannot be causal) as an approach to avoid the violation of the IV assumptions (such as through pleiotropy).^{33,34}

There are potential limitations to the current study. First, analyzing data sets (for the exposures and outcome) of individuals with the same ancestry

(i.e. European ancestry) helped reduce linkage disequilibrium, but our conclusions may have limited applications to individuals of non-European ancestry. Future studies can explore if there is a significant association between smoking traits and cSCC risk in populations of different ancestries. A previous single-center retrospective chart review study suggested a strong association between smoking and age of cSCC diagnosis in non-European populations,⁵⁵ and MR analysis can help further elucidate this possible association. Second, deriving exposures (ie smoking initiation and cigarettes per day) from GWAS summary data combining many heterogeneous cohorts (versus from a unique homogeneous cohort (i.e. UKB for lifetime smoking)) could be considered as a study limitation. Furthermore, cohort characteristics from which exposures were derived would have led to potential selection bias; for instance, lifetime smoking was derived from UKB, with participants being less likely to be a smoker and overall healthier than the general UK population.³⁶ Third, in the current study, we did not consider epigenetic changes that could modulate gene expression and distort the effect that MR finds between a genetic variant and an outcome. Thus, future MR studies ideally would incorporate information about epigenetic changes on the genetic variants in the data analysis.^{37,38} Fourth, we acknowledge the limitations of MR to help determine causal associations given the assumptions that must be made before applying this technique. By assuming that humans are all randomly assigned genetic variants from their parents, we have also assumed that these genes are fully penetrant and not significantly impacted by environmental factors such as UV exposure, diet, or other factors that may influence cSCC risk.^{4,39} In essence, disease development such as cSCC consists of multifactorial, dynamic, nonlinear biological processes, and MR is limited in accounting for that complexity. Future studies can expand on this research by analyzing the interplay between genetic predispositions for smoking, UV exposure, diet, and other behaviors on cSCC risk.

Although no causal associations of smoking initiation and smoking amount with SCC risk were detected, we observed support for a causal association of lifetime smoking with SCC risk. Lifetime smoking exposure is based on an index that incorporates total duration of smoking, time since smoking cessation, average number of cigarettes smoked per day, and selected covariable interactions to form an aggregate measurement.⁴⁰ For this reason, we feel that lifetime smoking might be a more comprehensive exposure for smoking behavior, compared to smoking initiation and amount smoked per day. Previous MR studies have reported causal

associations between lifetime smoking exposure and other cancers, including breast cancer and colorectal cancer, but a lack of association with other cigarette smoking traits.^{20,21} Thus, future MR studies that aim to investigate the causal relationship between cigarette smoking and outcomes should test for different smoking traits, including lifetime smoking index, as those represent different stages of cigarette use.

In this study analyzing GWAS summary statistics of different cohorts through two-sample MR, we found support for a causal relationship between long-term smoking exposure and cSCC risk. These results support more targeted screening strategies of populations based on factors such as carcinogen exposure and genetic and clinical information. Given that lifetime smoking exposure is an aggregate measurement incorporating total duration of smoking, time since smoking cessation, average number of cigarettes smoked per day, and selected covariable interactions, our results also support an overall reduction of smoking exposure. This conclusion may have broader implications in highlighting the importance of smoking cessation initiatives in decreasing smoking exposure. Causal relationships between smoking traits and a wide range of cancers, including cancers of the lung, head and neck, esophagus, pancreas, bladder, kidney, cervix, and ovaries, and myeloid leukemia, have been reported,²² and our results can potentially provide additional evidence for skin cancer. In conclusion, this study shows evidence for a causal relationship between long-term smoking exposure and cSCC risk, underscoring the importance of smoking prevention campaigns and more targeted screening strategies.

We are grateful to the Kaiser Permanente Northern California members who have generously agreed to participate in the Kaiser Permanente Research Program on Genes, Environment, and Health. Support for participant enrollment, survey completion, and biospecimen collection for the RPGEH was provided by the Robert Wood Johnson Foundation, the Wayne and Gladys Valley Foundation, the Ellison Medical Foundation, and Kaiser Permanente Community Benefit Programs. Genotyping of the GERA cohort was funded by a grant from the National Institute on Aging, National Institute of Mental Health, and National Institute of Health Common Fund (RC2 AG036607). This study was funded by the National Cancer Institute (NCI) grant R01 CA241623.

Access to data and data analysis: C. J. and H. C. had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Conflicts of interest

None disclosed.

REFERENCES

- Urban K, Mehrmal S, Uppal P, Giesey RL, Delost GR. The global burden of skin cancer: a longitudinal analysis from the Global Burden of Disease study, 1990-2017. *JAAD Int*. 2021;2:98-108. <https://doi.org/10.1016/j.jdin.2020.10.013>
- Fania L, Didona D, Di Pietro FR, et al. Cutaneous squamous cell carcinoma: from pathophysiology to novel therapeutic approaches. *Biomedicines*. 2021;9(2):171. <https://doi.org/10.3390/biomedicines9020171>
- Diseases GBD, Injuries C. Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease study 2019. *Lancet*. 2020;396(10258):1204-1222. [https://doi.org/10.1016/S0140-6736\(20\)30925-9](https://doi.org/10.1016/S0140-6736(20)30925-9)
- Choquet H, Ashrafzadeh S, Kim Y, Asgari MM, Jorgenson E. Genetic and environmental factors underlying keratinocyte carcinoma risk. *JCI Insight*. 2020;5(10):e134783. <https://doi.org/10.1172/jci.insight.134783>
- Walser T, Cui X, Yanagawa J, et al. Smoking and lung cancer: the role of inflammation. *Proc Am Thorac Soc*. 2008;5(8):811-815. <https://doi.org/10.1513/pats.200809-100TH>
- Jethwa AR, Khariwala SS. Tobacco-related carcinogenesis in head and neck cancer. *Cancer Metastasis Rev*. 2017;36(3):411-423. <https://doi.org/10.1007/s10555-017-9689-6>
- Jones ME, Schoemaker MJ, Wright LB, Ashworth A, Swerdlow AJ. Smoking and risk of breast cancer in the generations study cohort. *Breast Cancer Res*. 2017;19(1):118. <https://doi.org/10.1186/s13058-017-0908-4>
- Roura E, Castellsague X, Pawlita M, et al. Smoking as a major risk factor for cervical cancer and pre-cancer: results from the EPIC cohort. *Int J Cancer*. 2014;135(2):453-466. <https://doi.org/10.1002/ijc.28666>
- Lynch SM, Vrieling A, Lubin JH, et al. Cigarette smoking and pancreatic cancer: a pooled analysis from the pancreatic cancer cohort consortium. *Am J Epidemiol*. 2009;170(4):403-413. <https://doi.org/10.1093/aje/kwp134>
- Dusingize JC, Olsen CM, Pandeya NP, et al. Cigarette smoking and the risks of basal cell carcinoma and squamous cell carcinoma. *J Invest Dermatol*. 2017;137(8):1700-1708. <https://doi.org/10.1016/j.jid.2017.03.027>
- Pirie K, Beral V, Heath AK, et al. Heterogeneous relationships of squamous and basal cell carcinomas of the skin with smoking: the UK Million Women Study and meta-analysis of prospective studies. *Br J Cancer*. 2018;119(1):114-120. <https://doi.org/10.1038/s41416-018-0105-y>
- Song F, Qureshi AA, Gao X, Li T, Han J. Smoking and risk of skin cancer: a prospective analysis and a meta-analysis. *Int J Epidemiol*. 2012;41(6):1694-1705. <https://doi.org/10.1093/ije/dys146>
- Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet*. 2014;23(R1):R89-R98. <https://doi.org/10.1093/hmg/ddu328>
- Nitsch D, Molokhia M, Smeeth L, et al. Limits to causal inference based on Mendelian randomization: a comparison with randomized controlled trials. *Am J Epidemiol*. 2006;163(5):397-403. <https://doi.org/10.1093/aje/kwj062>
- Asgari MM, Wang W, Ioannidis NM, et al. Identification of susceptibility loci for cutaneous squamous cell carcinoma. *J Invest Dermatol*. 2016;136(5):930-937. <https://doi.org/10.1016/j.jid.2016.01.013>
- Kvale MN, Hesselton S, Hoffmann TJ, et al. Genotyping informatics and quality control for 100,000 subjects in the genetic epidemiology research on adult health and aging (GERA) cohort. Research support, N.I.H., extramural research support, non-U.S. Gov't *Genetics*. 2015;200(4):1051-1060. <https://doi.org/10.1534/genetics.115.178905>
- Banda Y, Kvale MN, Hoffmann TJ, et al. Characterizing race/ethnicity and genetic ancestry for 100,000 subjects in the genetic epidemiology research on adult health and aging (GERA) cohort. Research support, N.I.H., extramural research support, non-U.S. Gov't *Genetics*. 2015;200(4):1285-1295. <https://doi.org/10.1534/genetics.115.178616>
- Saunders GRB, Wang X, Chen F, et al. Genetic diversity fuels gene discovery for tobacco and alcohol use. *Nature*. 2022;612(7941):720-724. <https://doi.org/10.1038/s41586-022-05477-4>
- Choquet H, Yin J, Jorgenson E. Cigarette smoking behaviors and the importance of ethnicity and genetic ancestry. *Transl Psychiatry*. 2021;11(1):120. <https://doi.org/10.1038/s41398-021-01244-7>
- Wootton RE, Richmond RC, Stuijzand BG, et al. Evidence for causal effects of lifetime smoking on risk for depression and schizophrenia: a Mendelian randomisation study. *Psychol Med*. 2020;50(14):2435-2443. <https://doi.org/10.1017/S0033291719002678>
- Sudlow C, Gallacher J, Allen N, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med*. 2015;12(3):e1001779. <https://doi.org/10.1371/journal.pmed.1001779>
- Bulik-Sullivan BK, Loh PR, Finucane HK, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet*. 2015;47(3):291-295. <https://doi.org/10.1038/ng.3211>
- Hemani G, Zheng J, Elsworth B, et al. The MR-Base platform supports systematic causal inference across the human phenome. *Elife*. 2018;7:e34408. <https://doi.org/10.7554/eLife.34408>
- Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol*. 2015;44(2):512-525. <https://doi.org/10.1093/ije/dyv080>
- Burgess S, Davey Smith G, Davies NM, et al. Guidelines for performing Mendelian randomization investigations. *Wellcome Open Res*. 2019;4:186. <https://doi.org/10.12688/wellcomeopenres.15555.2>
- Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat Genet*. 2018;50(5):693-698. <https://doi.org/10.1038/s41588-018-0099-7>
- De Hertog SA, Wensveen CA, Bastiaens MT, et al. Relation between smoking and skin cancer. *J Clin Oncol*. 2001;19(1):231-238. <https://doi.org/10.1200/JCO.2001.19.1.231>
- Zheng J, Baird D, Borges MC, et al. Recent developments in mendelian randomization studies. *Curr Epidemiol Rep*. 2017;4(4):330-345. <https://doi.org/10.1007/s40471-017-0128-6>
- Davies NM, Holmes MV, Davey Smith G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. *Bmj*. 2018;362:k601. <https://doi.org/10.1136/bmj.k601>
- Qi G, Chatterjee N. A comprehensive evaluation of methods for Mendelian randomization using realistic simulations and an analysis of 38 biomarkers for risk of type 2 diabetes. *Int J Epidemiol*. 2021;50(4):1335-1349. <https://doi.org/10.1093/ije/dyaa262>
- Brown BC, Knowles DA. Welch-weighted Egger regression reduces false positives due to correlated pleiotropy in Mendelian randomization. *Am J Hum Genet*. 2021;108(12):2319-2335. <https://doi.org/10.1016/j.ajhg.2021.10.006>

32. Reed ZE, Wootton RE, Khouja JN, et al. Exploring pleiotropy in Mendelian randomisation analyses: what are genetic variants associated with “cigarette smoking initiation” really capturing? *medRxiv*. 2023;23293638. <https://doi.org/10.1101/2023.08.04.23293638>
33. Burgess S, Davey Smith G, Davies NM, et al. Guidelines for performing Mendelian randomization investigations: update for summer 2023. *Wellcome Open Res*. 2019;4:186. <https://doi.org/10.12688/wellcomeopenres.15555.3>
34. Sanderson E, Richardson TG, Hemani G, Davey Smith G. The use of negative control outcomes in Mendelian randomization to detect potential population stratification. *Int J Epidemiol*. 2021;50(4):1350-1361. <https://doi.org/10.1093/ije/dyaa288>
35. Nadhan KS, Chung CL, Buchanan EM, et al. Risk factors for keratinocyte carcinoma skin cancer in nonwhite individuals: a retrospective analysis. *J Am Acad Dermatol*. 2019;81(2):373-378. <https://doi.org/10.1016/j.jaad.2019.01.038>
36. Munafo MR, Tilling K, Taylor AE, Evans DM, Davey Smith G. Collider scope: when selection bias can substantially influence observed associations. *Int J Epidemiol*. 2018;47(1):226-235. <https://doi.org/10.1093/ije/dyx206>
37. Ogbuanu IU, Zhang H, Karmaus W. Can we apply the Mendelian randomization methodology without considering epigenetic effects? *Emerg Themes Epidemiol*. 2009;6:3. <https://doi.org/10.1186/1742-7622-6-3>
38. Relton CL, Davey Smith G. Two-step epigenetic Mendelian randomization: a strategy for establishing the causal role of epigenetic processes in pathways to disease. *Int J Epidemiol*. 2012;41(1):161-176. <https://doi.org/10.1093/ije/dyr233>
39. Lucock MD. The evolution of human skin pigmentation: a changing medley of vitamins, genetic variability, and UV radiation during human expansion. *Am J Biol Anthropol*. 2023;180(2):252-271. <https://doi.org/10.1002/ajpa.24564>
40. Leffondre K, Abrahamowicz M, Xiao Y, Siemiatycki J. Modelling smoking history using a comprehensive smoking index: application to lung cancer. *Stat Med*. 2006;25(24):4132-4146. <https://doi.org/10.1002/sim.2680>