OXFORD UNIVERSITY PRESS | International Journal of Epidemiology

Mendelian Randomization and mediation analysis of leukocyte telomere length and risk of lung and head and neck cancers

Journal:	*International Journal of Epidemiology*
Manuscript ID	IJE-2017-05-0616.R2
Manuscript Type:	Original Article
Date Submitted by the Author:	07-May-2018

Complete List of Authors:	Kachuri, Linda; Lunenfeld-Tanenbaum Research Institute; University of Toronto Dalla Lana School of Public Health Saarela, Olli; University of Toronto Dalla Lana School of Public Health Bojesen, Stig; Herlev Hospital, Department of Clinical Biochemistry Davey Smith, George; University of Bristol, Liu, Geoffrey; Ontario Cancer Institute, Toronto, ON, Canada, Division of Applied Molecular Oncology Landi, Maria Teresa; National Cancer Institute, Caporaso, Neil; National Cancer Institute, Christiani, David; Harvard School of Public Health, Environmental and Occupational Medicine and Epidemiology Program Johansson, Mattias; IARC, Panico, Salvatore; Universita degli Studi di Napoli Federico II Overvad, Kim; Aarhus University, Department of Public Health Trichopoulou, Antonia; WHO Collaborating Center for Food and Nutrition Policies, Department of Hygiene, Epidemiology and Medical Statistics Vineis, Paolo; imperial College London, Dept of Epidemiology & Public Health Scelo, Ghislaine; International Agency for Research on Cancer, Zaridze, David; Russian N.N.Blokhin Cancer Research Centre Wu, X; The University of Texas, Department of Epidemiology Albanes, Demetrius; National Cancer Institute, Division of Cancer Epidemiology and Genetics Diergaarde, Brenda; University of Pittsburgh, Department of Human Genetics, Graduate School of Public Health Lagiou, Pagona; University of Athens Medical School, Dept of Hygiene, Epidemiology Macfarlane, Gary; University of Aberdeen, Epidemiology Group Aldrich, Melinda; Vanderbilt University, Thoracic Surgery Tardon, Adonina; University of Oviedo, Molecular Epidemiology of Cancer Unit, University Institute of Oncology Rennert, Gad; Clalit National Cancer Control Center at Carmel Medical Center Olshan, Andrew; University of North Carolina at Chapel Hill Gillings School of Global Public Health Weissler, Mark; University of North Carolina at Chapel Hill, Department of

	Otolaryngology/Head and Neck Surgery Chen, Chu; Fred Hutchinson Cancer Research Center, Division of Public Health Sciences Goodman, Gary; Fred Hutchinson Cancer Research Center, Doherty, Jennifer; Dartmouth Medical School, Ness, A; University of Bristol, Department of Community Based Medicine Bickeböller, Heike; Georg-August-Universitat Gottingen Abteilung fur Molekulare Mikrobiologie und Genetik Wichmann, H.-Erich; Ludwig-Maximilians-University Munich, IBE-Chair of Epidemiology Risch, Angela; Deutsches Krebsforschungszentrum Field, John; University of Liverpool, Teare, M. Dawn; University of Sheffield, School of Health and Related Research Kiemeney, Lambertus A.L.M.; Radboud University Nijmegen Medical Centre, Van der Heijden, Erik; Radboud University Medical Centre Carroll, June; Mount Sinai Hospital Haugen, Aage; National Institute of Occupational Health Zienolddiny, Shanbeh; National Institute of Occupational Health Skaug, Vidar; National Institute of Occupational Health Wünsch-Filho, Victor; Faculty of Public Health of University of São Paulo, Epidemiology Tajara, Eloiza; Faculdade de Medicina de Sao Jose do Rio Preto, Department of Molecular Biology Ayoub Moysés, Raquel; Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo, Disciplina de Cirurgia de Cabeça e Pescoço Daumas Nunes, Fabio ; School of Dentistry, University of São Paulo, Department of Oral Pathology Lam, Stephen; British Columbia Cancer Agency Eluf-Neto, J.; Faculdade de Medicina da Universidade de São Paulo, Departamento de Medicina Preventiva Lacko, Martin; Maastricht University Medical Center, Department of Otorhinolaryngology, Head and Neck Surgery Peters, Wilbert; Radboud University Nijmegen Medical Center, Department of Gastroenterology LeMarchand, Loic; University of Hawaii Cancer Center, Duell, Eric; Catalan Institute of Oncology, Unit of Nutrition, Environment and Cancer Andrew, Angeline; Dartmouth College Geisel School of Medicine Franceschi, Silvia ; International Agency for Research on Cancer Schabath, Matthew; H Lee Moffitt Cancer Center and Research Institute Manjer, Jonas; Malmö University Hospital, Department of Surgery Arnold, Susanne; Markey Cancer Center Lazarus, Philip; Washington State University College of Pharmacy, Mukeriya, Anush ; Russian N.N.Blokhin Cancer Research Centre Swiatkowska, Beata; Instytut Medycyny Pracy im prof Jerzego Nofera Janout, Vladimír; Palacky University, Holcatova, Ivana; Institute of Public Health and Preventive Medicine, Second Faculty of Medicine, Charles University Stojsic, Jelena; Klinicki centar Srbije, Department of Thoracopulmonary Pathology, Service of Pathology Mates, Dana; National Institute of Public Health, Lissowska, Jolanta; M. Curie Memorial Cancer Center, Department of Cancer Epidemiology and Prevention Boccia, Stefania; Section of Hygiene - Institute of Public Health; Università Cattolica del Sacro Cuore, Fondazione Policlinico Universitario "Agostino Gemelli", L.go F. Vito, 1 – 00168 Rome, Italy. Lesseur, Corina; International Agency for Research on Cancer Zong, Xuchen; Lunenfeld-Tanenbaum Research Institute Mckay, James; International Agency for Research on Cancer

	Brennan, Paul; IARC, Amos, Christopher; Geisel School of Medicine, Dartmouth College, Center for Genomic Medicine Hung, Rayjean; Lunenfeld-Tanenbaum Research Institute
Key Words:	lung cancer, telomere length, chromosome 5p15.33, Mendelian randomization, mediation analysis, TERT

Mendelian Randomization and mediation analysis of leukocyte telomere length and risk of lung and head and neck cancers

Linda Kachuri[1,2], Olli Saarela[3], Stig Egil Bojesen[4,5], George Davey Smith[6], Geoffrey Liu[2,7] , Maria Teresa Landi[8], Neil E. Caporaso[8], David C. Christiani[9,10], Mattias Johansson[11], Salvatore Panico[12], Kim Overvad[13], Antonia Trichopoulou[14,15], Paolo Vineis[16], Ghislaine Scelo[11], David Zaridze[17], Xifeng Wu[18], Demetrius Albanes[8], Brenda Diergaarde[19], Pagona Lagiou[15], Gary J. Macfarlane[20], Melinda C. Aldrich[21], Adonina Tardón[22], Gad Rennert[23], Andrew F. Olshan[24], Mark C. Weissler[25], Chu Chen[26], Gary E. Goodman[26], Jennifer A. Doherty[27], Andrew R. Ness[28], Heike Bickeböller[29], H.-Erich Wichmann[30,31,32], Angela Risch[33], John K. Field[34], M. Dawn Teare[35], Lambertus A. Kiemeney[36], Erik H.F.M. van der Heijden[36], June C. Carroll[1], Aage Haugen[37], Shanbeh Zienolddiny[37], Vidar Skaug[37], Victor Wünsch-Filho[38], Eloiza H. Tajara[39], Raquel Ayoub Moysés[40], Fabio Daumas Nunes[41], Stephen Lam[42], Jose Eluf-Neto[43], Martin Lacko[44], Wilbert H. M. Peters[45], Loïc Le Marchand[46], Eric J. Duell[47], Angeline S. Andrew[48], Silvia Franceschi[11], Matthew B. Schabath[49], Jonas Manjer[50], Susanne Arnold[51], Philip Lazarus[52], Anush Mukeriya[17], Beata Swiatkowska[53], Vladimir Janout[54], Ivana Holcatova[55], Jelena Stojsic[56], Dana Mates[57], Jolanta Lissowska[58], Stefania Boccia[59], Corina Lesseur[11,60], Xuchen Zong[1], James D. McKay[11], Paul Brennan[11], Christopher I. Amos[61], Rayjean J. Hung*[1,2]

1. Lunenfeld-Tanenbaum Research Institute, Sinai Health System, Toronto, ON, Canada
2. Division of Epidemiology, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada
3. Division of Biostatistics, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada
4. Department of Clinical Biochemistry, Herlev and Gentofte Hospital, Copenhagen University Hospital, Denmark
5. Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
6. Population Health Science, Bristol Medical School, MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK
7. Ontario Cancer Institute, Princess Margaret Cancer Center, Toronto, ON, Canada
8. Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA
9. Departments of Epidemiology and Environmental Health, Harvard TH Chan School of Public Health, Boston, MA, USA
10. Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA
11. International Agency for Research on Cancer, Lyon, France
12. Dipartimento di Medicina Clinica e Chirurgia, Federico II University, Naples, Italy
13. Aarhus University, Department of Public Health, Section for Epidemiology, Aarhus, Denmark
14. Hellenic Health Foundation, and WHO Collaborating Center for Nutrition and Health, Unit of Nutritional Epidemiology and Nutrition in Public Health, Athens, Greece
15. Department of Hygiene, Epidemiology and Medical Statistics, School of Medicine, National and Kapodistrian University of Athens, Greece
16. MRC/PHE Centre for Environment and Health, Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, United Kingdom
17. Department of Epidemiology and Prevention, Russian N.N.Blokhin Cancer Research Centre, Moscow, Russian Federation

18. Department of Epidemiology, Division of Cancer Prevention and Population Sciences, The University of Texas MD Anderson Cancer Center, Houston, TX, USA
19. Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA
20. The Institute of Applied Health Sciences, School of Medicine, University of Aberdeen, Aberdeen, United Kingdom
21. Department of Thoracic Surgery and Division of Epidemiology, Vanderbilt University Medical Center, Nashville, TN, USA
22. University of Oviedo and CIBERESP, Faculty of Medicine, Campus del Cristo, Oviedo, Spain
23. Clalit National Cancer Control Center at Carmel Medical Center and Technion Faculty of Medicine, Haifa, Israel
24. Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA.
25. Department of Otolaryngology/Head and Neck Surgery, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
26. Fred Hutchinson Cancer Research Center, Seattle, WA, USA
27. Department of Population Health Sciences, Huntsman Cancer Institute, Salt Lake City, Utah, USA
28. School of Oral and Dental Sciences, University of Bristol, Bristol, UK.
29. Department of Genetic Epidemiology, University Medical Center, Georg-August-University, Göttingen, Germany
30. Institute of Epidemiology II, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany
31. Institute of Medical Informatics, Biometry and Epidemiology, Ludwig Maximilians University, Munich, Germany
32. Institute of Medical Statistics and Epidemiology, Technical University, Munich, Germany
33. Division of Epigenomics & Cancer Risk Factors, German Cancer Research Center (DKFZ), Heidelberg, Germany
34. Roy Castle Lung Cancer Research Programme, University of Liverpool Cancer Research Centre Institute of Translational Medicine, University of Liverpool, Liverpool, UK
35. School of Health and Related Research, University of Sheffield, Sheffield, UK
36. Radboud Institute for Health Sciences, Radboud University Medical Centre, Nijmegen, The Netherlands
37. The National Institute of Occupational Health, Oslo, Norway
38. Faculdade de Saúde Pública, Universidade de São Paulo, Brazil
39. Department of Molecular Biology, School of Medicine of São José do Rio Preto, São José do Rio Preto, Brazil
40. Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo, Disciplina de Cirurgia de Cabeça e Pescoço (LIM28), São Paulo, Brasil
41. Department of Oral Pathology, School of Dentistry, University of São Paulo, São Paulo, Brazil
42. BC Cancer Agency, Vancouver, BC, Canada
43. Faculdade de Medicina da Universidade de São Paulo, Departamento de Medicina Preventiva, São Paulo, Brazil
44. Department of Otorhinolaryngology, Head and Neck Surgery, Maastricht University Medical Center, Maastricht, The Netherlands
45. Department of Gastroenterology, Radboud University Nijmegen Medical Center, Nijmegen, The Netherlands
46. Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI, USA
47. Unit of Nutrition and Cancer, Catalan Institute of Oncology (ICO-IDIBELL), Barcelona, Spain
48. Department of Epidemiology, Geisel School of Medicine, Dartmouth College, Hanover, NH, USA
49. Department of Cancer Epidemiology, H. Lee Moffitt Cancer Center & Research Institute, Tampa, FL, USA
50. Skåne University Hospital, Lund University, Lund, Sweden
51. University of Kentucky, Markey Cancer Center, Lexington, KY, USA
52. College of Pharmacy, Washington State University, Spokane, WA, USA
53. Nofer Institute of Occupational Medicine, Department of Environmental Epidemiology, Lodz, Poland
54. Faculty of Health Sciences, Palacky University, Olomouc, Czech Republic

55. Institute of Public Health and Preventive Medicine, Charles University, Second Faculty of Medicine, Prague, Czech Republic
56. Department of Thoracopulmonary Pathology, Service of Pathology, Clinical Center of Serbia, Belgrade, Serbia
57. National Institute of Public Health, Bucharest, Romania
58. Department of Cancer Epidemiology and Prevention, Cancer Center Maria Sklodowska-Curie Institute of Oncology, Warsaw, Poland
59. Section of Hygiene, Institute of Public Health, Università Cattolica del Sacro Cuore, Fondazione Policlinico Agostino Gemelli, Rome, Italy
60. Icahn School of Medicine at Mount Sinai, New York, NY, USA
61. Institute for Clinical and Translational Research, Baylor College of Medicine, Houston, TX, USA

Type of manuscript: Original Research Article

Short title (50 characters max): Telomere length and risk of lung and head and neck cancers

Word count: Abstract (250 words) Main text (4018 words)

Corresponding author:

Rayjean J. Hung, Ph.D., M.S.

Lunenfeld-Tanenbaum Research Institute of Sinai Health System

Division of Epidemiology, Dalla Lana School of Public Health, University of Toronto

60 Murray St. Room L5-215, Box 18

Toronto ON M5T 3L9 Canada

Tel: (416) 586-4750; Fax: (416) 586-8404

E-mail: rayjean.hung@lunenfeld.ca

ABSTRACT

Background: Evidence from observational studies of telomere length (TL) has been conflicting regarding its direction of association with cancer risk. We investigated the causal relevance of TL for lung and head and neck cancers using Mendelian Randomization (MR) and mediation analyses.

Methods: We developed a novel genetic instrument for TL in chromosome 5p15.33, using variants identified through deep-sequencing, that were genotyped in 2051 cancer-free subjects. Next, we conducted an MR analysis of lung (16396 cases, 13013 controls) and head and neck cancer (4415 cases, 5013 controls) using 8 genetic instruments for TL. Lastly, the 5p15.33 instrument and distinct 5p15.33 lung cancer risk loci were evaluated using two-sample mediation analysis, to quantify their direct and indirect, telomere-mediated, effects.

Results: The multi-allelic 5p15.33 instrument explained 1.49-2.00% of TL variation in our data ($p=2.6\times10^{-9}$). The MR analysis estimated that a 1000 base pair increase in TL increases risk of lung cancer (OR=1.41, 95% CI: 1.20-1.65) and lung adenocarcinoma (OR=1.92, 95% CI: 1.51-2.22), but not squamous lung carcinoma (OR=1.04, 95% CI: 0.83-1.29), or head and neck cancers (OR=0.90, 95% CI: 0.70-1.05). Mediation analysis of the 5p15.33 instrument indicated an absence of direct effects on lung cancer risk (OR=1.00, 95% CI: 0.95-1.04). Analysis of distinct 5p15.33 susceptibility variants estimated that TL mediates up to 40% of the observed associations with lung cancer risk.

Conclusions: Our findings support a causal role for long telomeres in lung cancer etiology, particularly for adenocarcinoma, and demonstrate that telomere maintenance partially mediates the lung cancer susceptibility conferred by 5p15.33 loci.

KEY MESSAGES

- Genetic predisposition to long telomeres increases risk of lung cancer, predominately lung adenocarcinoma
- Genetic determinants of long telomeres are not associated with squamous carcinomas of the lung or head and neck
- Using two-sample mediation analysis we determined that the novel 5p15.33 instrument for telomere length does not have direct effects on the outcome, and demonstrated that the association between 5p15.33 lung cancer susceptibility variants is partially mediated by telomere length, suggesting the presence of other relevant mechanisms

INTRODUCTION

Telomeres are highly conserved stretches of tandem repeats of the TTAGGG sequence, which protect chromosome ends from degradation and maintain genome stability(1, 2). Due to the incomplete replication of chromosomes during cell division, human telomeres lose between 50 and 200 base pairs with each replication(1-3). In checkpoint proficient cells critically short telomeres trigger senescence, followed by apoptosis, which represents a barrier against cancer initiation by limiting cellular proliferation(4, 5). As telomeres shorten their ability to maintain chromosomal stability also diminishes, which may increase cancer susceptibility(6, 7). However, long telomeres may also promote cancer development through an accumulation of mutations due to prolonged cell survival and proliferation. In fact, cancer cells are characterized by such a proliferative advantage, often through reactivation of telomerase, which is normally silent in somatic cells(4, 5, 8).

Telomere length (TL) has been studied extensively in relation to cancer risk. However, findings of epidemiologic studies have been conflicting (6, 9-11). Observational studies investigating TL measured after cancer diagnosis are particularly vulnerable to reverse causation and residual confounding, therefore shorter TL observed in cancer cases is likely to reflect underlying disease or the impact of cancer treatment (12, 13). It is also difficult to isolate the influence of TL on cancer risk from that of other risk factors that influence both TL and cancer susceptibility, including biological or replicative age (10, 14, 15).

Mendelian Randomization (MR) is an approach for evaluating causality by using single nucleotide polymorphisms (SNPs) in relevant genes as instrumental variables (IVs) (16). Genome-wide association studies (GWAS) identified a number of genetic regions involved in TL regulation, including genes encoding the catalytic subunit of telomerase (*TERT*) in chromosome 5p15.33 and its RNA template (*TERC*) in 3q26.2 (17-21). By leveraging these associations, MR can provide a valid test of the causal hypothesis assuming the genetic IVs only affect cancer risk through TL regulation.

Previous studies using genetic proxies for TL suggest that longer telomeres confer an increased risk of lung cancer, especially adenocarcinoma (22-24), which is consistent with the findings of prospective observational studies (25-27). Lung cancer case-control studies report both increased (28) and inverse (6, 29) associations for long TL, and some implicate high TL variability in lung cancer susceptibility (30). For head and neck cancers (HNC), which are predominantly squamous carcinomas, short TL is consistently associated with increased risk in case-control studies (6, 31, 32), whereas a recent MR analysis (24) did find evidence supporting a causal relationship.

The overarching aim of this study is to investigate the causal relationship between TL and risk of lung and upper aero-digestive tract cancers. First, we developed a novel genetic instrument for TL in chromosome 5p15.33, given the extensive pleiotropy in this region and potential for violating MR assumptions (22, 33). Next, we conducted the largest two-sample MR analysis of lung and HNC risk to date. Lastly, we quantified the direct and telomere-mediated effects of 5p15.33 genetic variants on cancer risk using a two-sample mediation analysis approach (Figure 1).

METHODS

Study populations

We used individual-level data from 23 pooled studies of lung cancer, with 16396 cases (5690 adenocarcinoma, 4045 squamous carcinoma) and 13013 controls; and 11 HNC studies with 4415 cases and 5013 controls, all part of the OncoArray collaboration (34) (Supplementary Tables 1-2). Descriptions of studies and genotyping methods have been previously published (34, 35) (details in Supplementary File 1). Analyses were restricted to individuals of predominantly European ancestry (≥80% lung, >70% HNC)(34, 36). Studies received approval from institutional research ethics review boards and informed consent was obtained from the participants.

The novel 5p15.33 instrument was developed using data from two studies: the cancer-free controls from the Mount Sinai and Princess Margaret Hospital (MSH-PMH) case-control study in Toronto(37), and cancer-free individuals from the Copenhagen General Population Study

(CGPS)(38), a population-based prospective cohort (Table 1). TL was measured in DNA from peripheral blood leukocytes using previously described quantitative polymerase chain reaction assays performed in MSH-PMH (37) and CGPS (23, 38) (details in Supplementary File 2).

Statistical Analysis

Mendelian randomization analysis

The genetic instruments for TL included independent SNPs showing strong prior evidence of association with TL, such as $p<5\times10^{-8}$ in the discovery stage of at least one GWAS and replication in a separate GWAS or meta-analysis (17-21). In addition to the new 5p15.33 instrument described below, we selected 7 additional loci involved in telomere maintenance: rs10165485 (proxy for rs11125529, r^2=1.0) in *ACYP2* (2p16.2), rs6772228 in *PXK* (3p14.3), rs10936599 in *TERC* (3q26.2), rs11100479 (proxy for rs7675998, r^2=0.99) in *NAF1* (4q32.2), rs9420907 in *OBFC1* (10q24.3), rs10419926 in *ZNF676* (19p12), and rs755017 near *RTEL1* and *ZBTB46* (20q13). Only genotyped, non-imputed variants were used.

For the purpose of developing a new instrument in the 5p15.33 region, TL values were converted to Z-scores in MSH-PMH (n=879) and CGPS (n=1172) studies separately, and pooled to increase statistical power. Linear regression was used to estimate the association between 899 variants in 5p15.33 and TL, adjusting for age, sex, study, and the top 5 genetic ancestry principal components (PCs).

Selection of variants for the 5p15.33 instrument was based on statistical significance, consistency across the two studies, and instrument strength, measured by the *F* statistic, which depends on the variance in TL explained by the genetic predictors (R^2), sample size (n), and number of instruments (k): $F=\left(\frac{n-k-1}{k}\right)\left(\frac{R^2}{1-R^2}\right)$. Variants were considered for inclusion in the 5p15.33 instrument if they met the following criteria:

i. F≥5 and p<0.05 in the Toronto and Copenhagen combined dataset (n=2051)

ii. F<5 and p<0.05 overall (n=2051) and F>5 among never smokers (n=848)

iii. Consistent direction of allelic effects in MSH-PMH and CGPS

iv. Minor allele detected in at least 2 individuals

Independent genetic variants (r^2<0.2) that met the selection criteria were combined into an allele score representing the 5p15.33 region to increase the power of the resulting instrument (39, 40).

The MR analysis combined summary statistics across the genetic IVs to estimate the causal parameter β_{IV}, which is the log odds ratio (OR) describing the causal effect of increasing TL on cancer risk (Supplementary Figure 1). Parameters for the MR analysis included β_{TL} and β_Y, where β_{TL} is a vector of SNP-TL associations and β_Y is a vector of per-allele cancer log ORs for each instrument. For genetic instruments outside of 5p15.33, β_{TL} and corresponding standard errors (SE) were obtained from the literature and scaled to represent a 1000 base pair (kbp) increase in leukocyte TL, a proxy for TL in relevant tissues(19-21). For all instruments, β_Y and corresponding SE were estimated directly using individual-level OncoArray lung and HNC data. Logistic regression models were adjusted for age, sex, study, and 10 PCs.

The causal parameter β_{IV} was estimated using the maximum likelihood-based (ML) approach and the inverse-variance weighted (IVW) method (41, 42). This was complemented by sensitivity analyses using the weighted median estimator (WME), which provides valid estimates of the causal parameter even when up to 50% of the statistical weights are contributed by genetic instruments violate MR assumptions (43).

Mediation analysis

The aim of the mediation analysis was to quantify how much of the lung cancer association in the 5p15.33 region is mediated by TL. First, we validated the 5p15.33 instrument by decomposing its total effect on lung cancer into direct and indirect effects, mediated by TL. Next, we

extended this analysis to independent (r^2<0.20) variants that capture the lung cancer association signal in 5p15.33 (details in Supplementary File 3).

Our mediation approach is based on the counterfactual framework(44, 45) and extends the sensitivity analysis using two randomized controlled trials proposed by Vanderweele, which allows the mediator-outcome (θ_2) and exposure-mediator (β_1) relationships to be estimated in separate studies (46). Application of this approach in the present context assumes that a valid estimate for the mediator-outcome relationship can be obtained from an independent MR or cohort studies. Based on previously published formulas for mediation analysis (44, 45), the total effect (TE) of increasing the exposure from reference level a^* to level a on lung cancer (Y) conditional on covariates *c* can be decomposed into natural direct effects (NDE) and natural indirect effects (NIE):

$$OR^{TE}_{a,a^*|c} = \frac{P(Y_a = 1 \mid c) / \{1 - P(Y_a = 1 \mid c)}{P(Y_{a^*} = 1 \mid c) / \{1 - P(Y_{a^*} = 1 \mid c)} = OR^{NIE}_{a,a^*|c} \times OR^{NDE}_{a,a^*|c} \quad (1)$$

Assuming a rare outcome and absence of exposure-mediator interaction, mediated effects are given by:

$$OR^{NIE}_{a,a^*|c} \approx \exp\{\theta_2 \times \beta_1 (a - a^*)\} \quad (2)$$

where θ_2 is log-OR per one unit increment in TL and β_1 is the effect of the 5p15.33 instrument on TL. Based on equation 1, NDE can be obtained by subtracting the NIE from the total effect:

$$\log(OR^{NDE}_{a,a^*|c}) \approx \log(OR^{TE}_{a,a^*|c}) - \log(OR^{NIE}_{a,a^*|c}) \quad (3)$$

In the presence of interaction between the exposure and mediator, the NIE is given by:

$$OR^{NIE}_{a,a^*|c} \approx \exp\{(\theta_2 \times \beta_1 + \theta_3 \times \beta_1 a) \times (a - a^*)\} \quad (4)$$

where θ_2 now represents the main effect of the mediator, TL, and θ_3 is the exposure-mediator interaction parameter, with NDE having a more complicated form given by Valeri and VanderWeele(45). Formulas for a dichotomized mediator are provided in Supplementary File 4.

The β_1 parameter for the 5p15.33 instrument is equivalent to β_{TL} estimated in the cancer-free subset of the MSH-PMH and CGPS studies, adjusting for appropriate covariates. For 5p15.33 cancer susceptibility variants, β_1 estimates were selected from Bojesen et al. (47), the largest fine-mapping analysis of common 5p15.33 loci and TL with 15567 cancer-free controls. Per allele associations were reported as percent increase in TL and base-pair change. OR^{TE} for all variants was estimated in 23 lung cancer OncoArray studies, and is equivalent to β_Y for the 5p15.33 instrument.

External estimates of the mediator-outcome relationship (θ_2) were substituted into the equation (2) to avoid estimating the effect of TL on lung cancer risk directly using MSH-PMH case-control data, which are likely to be biased due to the post-diagnostic timing of TL measurement. The effect of TL on lung cancer risk was obtained from two studies: an MR analysis TL by Zhang et al.(22), and a meta-analysis of prospective studies by Zhu et al. (11) (Supplementary Figure 2).

Since interaction between the 5p15.33 instrument and TL is plausible, we conducted sensitivity analyses under different magnitudes of θ_3 (details in Supplementary File 4). Confidence intervals for the NIE and NDE were approximated as Bayesian credible intervals. Analyses were conducted using R version 3.3.3.

RESULTS

Characteristics of the combined Toronto and Copenhagen dataset (n=2051), used to develop the 5p15.33 instrument, are summarized in Table 1. The cancer-free participants in the MSH-PMH and CGPS studies were of similar mean age, 61.0 and 61.30 years, respectively. Age was the strongest predictor of TL ($p=2.6\times10^{-30}$), while sex, smoking status, and cigarette pack-years among smokers were not associated with relative TL (Supplementary Table 3).

Novel 5p15.33 instrument for telomere length

The 5p15.33 variants comprising this instrument were not used in any previous MR studies of TL. After excluding 17 singletons and other SNPs that did not meet our criteria, 14 variants were

included in the multi-allelic instrument for 5p15.33 (Table 2; regional plot and LD illustrated in Supplementary Figure 3). Most variants were located in non-coding intronic regions of several genes, including *SLC6A3*, *TERT*, *LPCAT1*, and a long-noncoding RNA (*LINC01511*) except for rs35033501, a synonymous *TERT* variant. The resulting multi-allelic 5p15.33 IV accounted for 1.49% of variation in the telomere Z-score in all subjects (F = 35.83; β_{TL}= 0.14, SE=0.02) and 2.00% in never smokers (F = 20.81), but was not predictive of smoking status (p=0.19) or cigarette pack-years among smokers (p=0.59) (Table 3). The 5p15.33 instrument was positively associated with lung cancer (OR=1.04, 95% CI: 1.01-1.07) and lung adenocarcinoma (OR=1.06, 1.03-1.10), but not squamous lung carcinomas (OR=1.03, 0.98-1.07). An inverse association was observed for HNC (OR=0.95, 0.90-1.00) and oral cavity cancer (OR=0.93, 0.87-0.98).

Telomere length and cancer risk

Results of the MR analysis based on 8 genetic instruments are presented in Table 4 and Figure 2. The likelihood-based model estimated a 41% increase in lung cancer risk per kbp increase in TL (OR_{ML}=1.41, 95% CI: 1.20-1.65). Estimates of the causal OR for lung cancer remained consistent across MR estimation methods. Genetic determinants of TL were predominantly associated with adenocarcinoma (OR_{ML}=1.92, 1.51-2.45), and appeared unrelated to squamous carcinoma (OR_{ML}=1.04, 0.83-1.29) and small cell carcinoma (OR_{ML}=1.03, 0.76-1.39).

The effect of long TL on lung cancer risk was larger in magnitude among never smokers (OR_{ML}=1.78, 1.22-2.61) compared to smokers (OR_{ML}=1.36, 1.14-1.63), although the former was attenuated in sensitivity analyses (OR_{WME}=1.55, 95% CI: 0.98-2.46). Effects on adenocarcinoma risk were also substantial in never smokers (OR_{ML}=2.68, 1.70-4.24). Genetic determinants of long telomeres conferred a 68% increase in lung cancer risk (OR_{ML}=1.68, 1.07-2.62) in subjects aged 50 years or younger. In contrast to lung cancer, genetic predisposition for longer TL did not seem related to risk of HNC overall (OR_{ML}= 0.90, 0.70-1.05), oral cavity (OR_{ML}=0.88, 0.65-1.19) and oropharynx cancers (OR_{ML}=0.83, 0.59-1.16).

Several additional sensitivity analyses were undertaken to further interrogate the MR results. Since smoking is an established risk factor for both HNC and lung cancer, MR analyses were repeated with adjustment for cigarette pack-years and smoking status. No appreciable changes were observed in the causal effect estimates for lung cancer overall (OR_{ML}=1.50, 1.27-1.78), lung adenocarcinoma (OR_{ML}=1.95, 1.53-2.49), HNC (OR_{ML}=0.91, 0.67-1.23), oral cavity (OR_{ML}=0.82, 0.57-1.18) or oropharynx cancers (OR_{ML}=0.86, 0.57-1.31).

The potential for directional pleiotropy was evaluated by checking for asymmetry in the plots depicting ratio estimates for each instrument, β_Y / β_{TL}, plotted against instrument strength, $\beta_{TL} / SE(\beta_Y)$ (Supplementary Figure 4). These results were not suggestive of pleiotropy and none of the genetic instruments were associated with cigarette smoking status or pack-years (Supplementary Table 4). Lastly, selected causal effects were re-estimated using the weighted mode-based estimator (MBE), which is robust to horizontal pleiotropy when the largest number of similar causal effect estimates are based on valid instruments, even if the majority of instruments are invalid (48). Estimates for lung cancer overall (OR_{MBE}=1.34, 1.08-1.66), lung adenocarcinoma (OR_{MBE}=1.55, 1.14-2.12), and adenocarcinoma in never smokers (OR_{MBE}=2.04, 1.04-4.04), were consistent with the primary results in Table 4.

Mediation analysis of the 5p15.33 instrument

We conducted mediation analyses to quantify direct (OR^{NDE}) and indirect effects (OR^{NIE}) of the 5p15.33 instrument on lung cancer. The OR^{NIE} we report is the proportional change in the odds of lung cancer for a change in TL that occurs when the 5p15.33 allele score increases by one from the reference level, corresponding to the mean of the allele score distribution. The estimate of the TL effect on lung cancer (θ_2) was selected from the strict model reported by Zhang et al.(22) (OR per kbp increase: 1.37, 95% CI: 1.12-1.68), which excluded rs2736100 (*TERT*). OR^{TE} for the 5p15.33 IV was re-estimated after removing overlapping subjects (n=3498) between the OncoArray and Zhang et al.(22). Assuming no interaction between the 5p15.33 IV and TL, the lung cancer effect appeared to be almost entirely mediated by TL (OR^{NIE}=1.05, 1.01-1.08), whereas the direct

effects of the 5p15.33 IV appeared null (OR^{NDE}=1.00, 0.95-1.04) (Figure 3; Supplementary Table 5). For lung adenocarcinoma, the 5p15.33 effects mediated by TL were larger in magnitude (OR^{NIE}=1.11, 1.05-1.18) than direct effects, which were close to unity (OR^{NDE}=0.97, 0.90-1.03).

Interaction sensitivity analyses for the NIE and NDE were carried out across three levels of θ_3: 0.10, 0.20 and 0.30. As the magnitude of the interaction parameter increased, so did the NIE, while TL-independent effects were not observed (Figure 3). Indirect effects on lung cancer risk mediated by TL ranged from OR^{NIE}=1.06 (95% CI: 1.03-1.10) for θ_3=0.10, to OR^{NIE}=1.09 (95% CI: 1.05-1.15) for θ_3= 0.30. For adenocarcinoma, increasing the magnitude of interaction between the 5p15.33 IV and TL was also associated with increasing NIE and diminishing direct effects.

The prospective meta-analysis estimate of θ_2 from Zhu et al.(11) reported an OR of 1.28 (95% CI: 1.09-1.50) for lung cancer comparing long vs. short TL. Based on this binary mediator, the NIE mediated by TL was attenuated, but remained statistically significant (OR^{NIE}=1.01, 1.00-1.03). A positive direct effect on lung cancer risk was also observed (OR^{NDE}=1.03, 1.00-1.06). Assuming interaction between the 5p15.33 instrument and TL, the mediated effects ranged from OR^{NIE}=1.02 (95% CI: 1.01-1.03) when θ_3=0.10, to OR^{NIE}=1.03 (95% CI: 1.01-1.05) when θ_3=0.30, while the direct effects decreased (Figure 3; Supplementary Table 5).

Mediation analysis of 5p15.33 lung cancer susceptibility loci

Five common (MAF>0.05), independent (r^2 <0.20) variants were selected to represent the lung cancer susceptibility signal in 5p15.33 (details in Supplementary File 3): rs7705526 ($P_{Adeno}=4.6\times10^{-13}$; $P_{Lung}=8.0\times10^{-7}$), rs2736108 ($P_{Adeno}=1.7\times10^{-12}$; $P_{Lung}=1.8\times10^{-11}$), rs421629 ($P_{Adeno}=6.2\times10^{-9}$; $P_{Lung}=1.2\times10^{-16}$), rs13167280 ($P_{Adeno}=1.4\times10^{-8}$; $P_{Lung}=1.1\times10^{-6}$), and rs56345976 ($P_{Adeno}=2.2\times10^{-7}$; $P_{Lung}=3.6\times10^{-9}$). These variants have been associated with lung cancer and lung adenocarcinoma in previous studies (37, 49-51), and are representative of the genetic susceptibility architecture in this region.

Estimates of β_1 were obtained from Bojesen et al.(47), and three *TERT* lung cancer risk variants were significantly associated with TL: rs7705526 (P_{TL}=2.3×10^{-14}), rs2736108 (P_{TL}=5.8×10^{-7}), and rs13167280 (P_{TL}=1.2×10^{-5}). Estimates of θ_2 were selected from the MR analysis (22) and OR^{TE} were re-estimated for each variant after removing the overlapping subjects. For all variants, the TL-increasing allele was positively associated with cancer risk, and both direct and indirect, TL-mediated effects were significant (Supplementary Table 6).

For lung cancer, the proportion mediated (PM) by TL was the largest for rs13167280 (OR^{NIE}=1.05, 1.03-1.07; PM=40.5%), followed by rs7705526 (OR^{NIE}=1.03, 1.01-1.05; PM=28.7%) and rs2736108 (OR^{NIE} 1.02, 1.01-1.03; PM=13.7%). The magnitude and proportion of the SNP effects that were mediated by TL were larger for adenocarcinoma compared to lung cancer overall: rs7705526 (OR^{NIE}=1.07, 1.04-1.10; PM=36.5%), rs13167280 (OR^{NIE}=1.05, 1.03-1.07; PM=24.8%), and rs2736108 (OR^{NIE}=1.04, 1.03-1.06; PM=22.9%).

DISCUSSION

We observed an association between genetic determinants of long telomeres and increased risk of lung, but not head and neck cancers. Our findings lend support to a causal relationship between longer leukocyte TL and increased risk of lung adenocarcinoma, but not squamous or small cell carcinoma. The magnitude of the increased risk was larger in never smokers and participants aged 50 or younger, consistent with a stronger influence of genetic susceptibility in individuals with a lower burden of modifiable risk factors (52). Although histology and smoking status are closely linked, our results suggest that the associations were histology-specific for adenocarcinoma (53, 54). Lastly, our mediation analysis demonstrated that mechanisms resulting in long telomeres mediate a proportion of the increase in lung cancer and lung adenocarcinoma risk conferred by 5p15.33 loci, and that the proportion of genetic susceptibility attributed to telomere maintenance differs between distinct 5p15.33 susceptibility loci.

Other analyses using multi-SNP telomere scores have also observed excess risks of lung cancer(22-24) and lung adenocarcinoma(22, 24), but did not observe an effect of TL on oral cancer

risk (23, 24). Opposite directions of effect for the 5p15.33 instrument on lung and HNC are consistent with earlier reports of opposing allelic effects for 5p15.33 SNPs on lung and oral cancer, respectively (35, 55). Leukocyte TL and functional *TERT* variants were previously reported to be unrelated to squamous HNC risk(56), although one study linked short TL to increased HNC risk based on rs2736100, which may be an invalid instrument(22, 57). With the exception of the 5p15.33 IV, the instruments used in this study overlap with those used in other MR analyses of TL (22-24).

Our findings lend support to the hypothesis that a greater number of telomere-increasing alleles increase lung cancer susceptibility. Although the precise molecular mechanisms remain to be elucidated, telomere maintenance may promote carcinogenesis by enabling prolonged cell survival and accumulation of mutations. This is supported by the hallmark observation that telomerase is overexpressed in 85-90% of adult tumors(8, 58), as well as recent data showing that long telomeres increase chromosomal instability(59) and promote immortalization of cancer cells(60). Excessively long telomeres may also be more fragile and dysfunctional, which is supported by the observation that *TERT* not only replenishes telomeres, but also regulates a trimming process to maintain TL homeostasis (61-63).

Differences in the effect of TL persisted after stratifying by smoking status, suggesting that underlying mechanisms differ across tissues and histological types. Longer TL does not appear to increase risk of small cell lung cancer or squamous lung carcinoma, the histology that also comprises 90% of HNC tumours, and for which the causal effect of tobacco smoking is the strongest(64). Since our genetic instruments are unrelated to smoking, confounding is unlikely to account for these differences. It is plausible that genetic predisposition for telomere maintenance offers some protection against genomic instability due to oxidative stress, declining regenerative capacity and immune function(7, 65, 66). Although human papillomavirus (HPV), a known cause of oropharynx cancer(67), has been reported to correlate with TL(31), the similarity of associations observed for oropharynx and oral cancers, only 2% of which are attributed to HPV(68), suggests that HPV infection is unlikely to modify the influence of TL.

This analysis has several important strengths. Genetic instruments represent are unaffected by reverse causality and are more likely to reflect causality due to the independence of genotypes from confounding factors. In addition to the large sample size, our analysis leveraged rich genetic data in 5p15.33, including rare sequence variations, to develop a robust, novel instrument. Furthermore, the use of multiple genetic instruments from essential genes for telomere maintenance mitigates the possibility for weak instruments bias and genetic confounding due to pleiotropy. The association between genetic predisposition to long TL and increased lung cancer risk persisted in analyses using the weighted median and mode-based estimators, which further supports the causal interpretation of these results.

Our mediation analysis offers insight not only by validating the new 5p15.33 instrument, by demonstrating an absence of direct effects, but also by formally quantifying the contribution of telomere-related mechanisms to the observed association between the established lung and adenocarcinoma susceptibility loci and lung cancer risk in this region. Although we confirmed that TL is an important molecular mechanism underlying the associations observed for 5p15.33 lung cancer risk loci, our results also indicated that only a fraction of these genetic effects operate through telomere maintenance. For instance, only 3-8% of the total effect of rs421629 (*CLPTM1L*) was mediated TL, and approximately half of the association between the *TERT* loci and lung cancer risk can be attributed to telomere mechanisms.

These findings are consistent with our knowledge that 5p15.33 is a complex susceptibility locus for multiple cancers(33, 55, 69) and GWAS peaks in this region also encompass non-cancer traits, such as red blood cell counts, prostate-specific antigen levels, and lung diseases(69-72). In addition, non-canonical functions of *TERT*, related to proliferation and differentiation via regulation of Wnt/β-catenin and Myc signaling, have been proposed(73). Therefore, although telomere maintenance is clearly an important 5p15.33 mechanism, cancer susceptibility loci in this region likely invoke additional pathways.

Several limitations of this work should be acknowledged. The time lag between genotype assignment at conception and the assessment of genetic effects on TL and cancer risk, as well as

the time-varying nature of TL, pose challenges for interpreting MR estimates of the causal effect (74). However, while genetic instruments do not recapitulate all aspects of telomere function and dynamics, they can still provide a valid test of the causal hypothesis that inherited predisposition to telomere maintenance increases lung cancer susceptibility (75). Secondly, genetic instruments for leukocyte TL may not be accurate proxies for TL in target tissues, which would reduce the power of our genetic instruments. However, the validity of instruments based on leukocyte TL is supported by correlation between TL in leukocytes and other tissues, including lung, and comparable rates of telomere shortening across somatic tissues (76-78). Thirdly, our MR analysis may be affected by winner's curse, with the magnitude and strength of association with TL observed in the discovery dataset likely to be exaggerated, particularly the 5p15.33 instrument. However, since the instrument discovery and MR analysis populations are independent, any potential bias in the causal parameter due to winner's curse or limited instrument strength will be towards the null (79). A related concern involves our ability to detect subtle effects of TL on cancer risk due to the modest proportion of variation in TL explained by our genetic instruments (approximately 5%), which is comparable to most genetic instruments for complex phenotypes (80-82). Based on our power calculations, this analysis was adequately powered (>80%) to detect effects with OR of 1.5 and above for all lung and HNC histological subtypes and smoking-stratified analyses.

Lastly, the validity of our mediation analysis depends in part on the validity of the published estimates of the mediator-outcome relationship. MR-based estimates of the mediator-outcome relationship are likely to satisfy the assumption of no unmeasured confounding, but must assume that all instruments used in Zhang et al. (22) were valid. While observational studies are more susceptible to confounding and bias due measurement error in the molecular mediator (83), a synthesis of prospective studies provides complementary evidence that does not depend on MR assumptions, and is less vulnerable to reverse causation than case-control designs.

In summary, we demonstrated that genetic determinants of long telomeres are associated with an increased risk of lung cancer, particularly adenocarcinoma. The associations observed for HNC were less consistent with a causal relationship, however we cannot preclude the possibility of

a very subtle telomere effects (OR<1.5). Using mediation analysis that incorporates independent published data, we validated the novel 5p15.33 instrument and quantified the proportion of the lung cancer association signal in 5p15.33 that is mediated by TL. While this work provides insight into the role of TL in cancer etiology, further research is needed to identify appropriate ways of utilizing this complex biomarker in the context of disease prevention or clinical intervention.

ACKNOWLEDGMENTS

The authors would like to acknowledge all of the participants involved in this research and the funders and support.

Linda Kachuri is a fellow in the Canadian Institutes of Health Research (CIHR) Strategic Training in Advanced Genetic Epidemiology (STAGE) program and is supported by the CIHR Doctoral Research Award from the Frederick Banting and Charles Best Canada Graduate Scholarships (GSD-137441).

Transdisciplinary Research for Cancer in Lung (TRICL) of the International Lung Cancer Consortium (ILCCO) was supported by the National Institutes of Health (U19-CA148127, CA148127S1). Genotyping for the TRICL-ILCCO OncoArray was supported by in kind genotyping at Centre for Inherited Disease Research (CIDR) (26820120008i-0-6800068-1). Genotyping for the Head and Neck Cancer OncoArray performed at CIDR was funded by the US National Institute of Dental and Craniofacial Research (NIDCR) grant 1X01HG007780-0.

CAPUA study was supported by FIS-FEDER/Spain grant numbers FIS-01/310, FIS-PI03-0365, and FIS-07-BI060604, FICYT/Asturias grant numbers FICYT PB02-67 and FICYT IB09-133, and the University Institute of Oncology (IUOPA), of the University of Oviedo and the Ciber de Epidemiologia y Salud Pública. CIBERESP, SPAIN.

The work performed in the CARET study was supported by the National Institute of Health (NIH) / National Cancer Institute (NCI): UM1 CA167462 (PI: Goodman), National Institute of Health UO1-CA6367307 (PIs Omen, Goodman); National Institute of Health R01 CA111703 (PI Chen), National Institute of Health 5R01 CA151989 (PI Doherty).

The Liverpool Lung Project is supported by the Roy Castle Lung Cancer Foundation.

The Harvard Lung Cancer Study was supported by the NIH (National Cancer Institute) grants CA092824, CA090578, CA074386.

The Multiethnic Cohort Study was partially supported by NIH Grants CA164973, CA033619, CA63464 and CA148127

The work performed in MSH-PMH study was supported by The Canadian Cancer Society Research Institute (020214), Ontario Institute of Cancer and Cancer Care Ontario Chair Award to R.J.H. and G.L. and the Alan Brown Chair and Lusi Wong Programs at the Princess Margaret Hospital Foundation.

The Norway study was supported by Norwegian Cancer Society, Norwegian Research Council

The work in TLC study has been supported in part the James & Esther King Biomedical Research Program (09KN-15), National Institutes of Health Specialized Programs of Research Excellence (SPORE) Grant (P50 CA119997), and by a Cancer Center Support Grant (CCSG) at the H. Lee

Moffitt Cancer Center and Research Institute, an NCI designated Comprehensive Cancer Center (grant number P30-CA76292)

The dataset(s) used for the analyses described were obtained from Vanderbilt University Medical Center's BioVU, which is supported by institutional funding and by the Vanderbilt CTSA grant UL1 TR000445 from NCATS/NIH. Dr. Melinda Aldrich is supported by the by NIH/National Cancer Institute 5K07CA172294.

The Copenhagen General Population Study (CGPS) was supported by the Chief Physician Johan Boserup and Lise Boserup Fund, the Danish Medical Research Council and Herlev Hospital.

The NELCS study: Grant Number P20RR018787 from the National Center for Research Resources (NCRR), a component of the National Institutes of Health (NIH).

Kentucky Lung Cancer Research Initiative (KLCRI) was supported by the Department of Defense [Congressionally Directed Medical Research Program, U.S. Army Medical Research and Materiel Command Program] under award number: 10153006 (W81XWH-11-1-0781). Views and opinions of, and endorsements by the author(s) do not reflect those of the US Army or the Department of Defense. This research was also supported by unrestricted infrastructure funds from the UK Center for Clinical and Translational Science, NIH grant UL1TR000117 and Markey Cancer Center NCI Cancer Center Support Grant (P30 CA177558) Shared Resource Facilities: Cancer Research Informatics, Biospecimen and Tissue Procurement, and Biostatistics and Bioinformatics.

The research undertaken by M.D.T., L.V.W. and M.S.A. was partly funded by the National Institute for Health Research (NIHR). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. M.D.T. holds a Medical Research Council Senior Clinical Fellowship (G0902313).

The Tampa study was funded by Public Health Service grants P01-CA68384 and R01-DE13158 from the National Institutes of Health.

The University of Pittsburgh head and neck cancer case–control study is supported by US National Institutes of Health grants P50 CA097190 and P30 CA047904.

The Carolina Head and Neck Cancer Study (CHANCE) was supported by the National Cancer Institute (R01CA90731).

The Head and Neck Genome Project (GENCAPO) was supported by the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP; grants 04/12054-9 and 10/51168-0). The authors thank all the members of the GENCAPO team.

This publication presents data from the Head and Neck 5000 study. The study was a component of independent research funded by the National Institute for Health Research (NIHR) under its Programme Grants for Applied Research scheme (RP-PG-0707-10034). The views expressed in

this publication are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. Human papillomavirus (HPV) serology was supported by a Cancer Research UK Programme Grant, the Integrative Cancer Epidemiology Programme (grant number: C18281/A19169).

The Alcohol-Related Cancers and Genetic Susceptibility Study in Europe (ARCAGE) was funded by the European Commission's fifth framework programme (QLK1- 2001-00182), the Italian Association for Cancer Research, Compagnia di San Paolo/FIRMS, Region Piemonte and Padova University (CPDA057222).

The Rome Study was supported by the Associazione Italiana per la Ricerca sul Cancro (AIRC) awards IG 2011 10491 and IG 2013 14220 to S.B. and by Fondazione Veronesi to S.B.

The IARC Latin American study was funded by the European Commission INCO-DC programme (IC18-CT97-0222), with additional funding from Fondo para la Investigación Científica y Tecnológica (Argentina) and the Fundação de Amparo à Pesquisa do Estado de São Paulo (01/01768-2).

The IARC Central Europe study was supported by the European Commission's INCO-COPERNICUS Program (IC15-CT98-0332), US NIH/National Cancer Institute grant CA92039 and World Cancer Research Foundation grant WCRF 99A28.

The IARC Oral Cancer Multicenter study was funded by grant S06 96 202489 05F02 from Europe against Cancer; grants FIS 97/0024, FIS 97/0662 and BAE 01/5013 from Fondo de Investigaciones Sanitarias, Spain; the UICC Yamagiwa-Yoshida Memorial International Cancer Study; the National Cancer Institute of Canada; Associazione Italiana per la Ricerca sul Cancro; and the Pan-American Health Organization. Coordination of the EPIC study is financially supported by the European Commission (DG SANCO) and the International Agency for Research on Cancer.

References

1. Blackburn EH. Structure and function of telomeres. Nature. 1991;350(6319):569-73.

2. de Lange T. Protection of mammalian telomeres. Oncogene. 2002;21(4):532-40.

3. Zhao Y, Sfeir AJ, Zou Y, Buseman CM, Chow TT, Shay JW, et al. Telomere extension occurs at most chromosome ends and is uncoupled from fill-in in human cancer cells. Cell. 2009;138(3):463-75.

4. Bodnar AG, Ouellette M, Frolkis M, Holt SE, Chiu CP, Morin GB, et al. Extension of life-span by introduction of telomerase into normal human cells. Science. 1998;279(5349):349-52.

5. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144(5):646-74.

6. Wu X, Amos CI, Zhu Y, Zhao H, Grossman BH, Shay JW, et al. Telomere dysfunction: a potential cancer predisposition factor. J Natl Cancer Inst. 2003;95(16):1211-8.

7. Bernardes de Jesus B, Blasco MA. Telomerase at the intersection of cancer and aging. Trends in genetics : TIG. 2013;29(9):513-20.

8. Newbold RF. The significance of telomerase activation and cellular immortalization in human cancer. Mutagenesis. 2002;17(6):539-50.

9. Wentzensen IM, Mirabello L, Pfeiffer RM, Savage SA. The association of telomere length and cancer: a meta-analysis. Cancer Epidemiol Biomarkers Prev. 2011;20(6):1238-50.

10. Prescott J, Wentzensen IM, Savage SA, De Vivo I. Epidemiologic evidence for a role of telomere dysfunction in cancer etiology. Mutat Res. 2012;730(1-2):75-84.

11. Zhu X, Han W, Xue W, Zou Y, Xie C, Du J, et al. The association between telomere length and cancer risk in population studies. Scientific reports. 2016;6:22243.

12. Benitez-Buelga C, Sanchez-Barroso L, Gallardo M, Apellaniz-Ruiz M, Inglada-Perez L, Yanowski K, et al. Impact of chemotherapy on telomere length in sporadic and familial breast cancer patients. Breast Cancer Res Treat. 2015;149(2):385-94.

13. Li P, Hou M, Lou F, Bjorkholm M, Xu D. Telomere dysfunction induced by chemotherapeutic agents and radiation in normal human cells. Int J Biochem Cell Biol. 2012;44(9):1531-40.

14. Huzen J, Wong LS, van Veldhuisen DJ, Samani NJ, Zwinderman AH, Codd V, et al. Telomere length loss due to smoking and metabolic traits. J Intern Med. 2014;275(2):155-63.

15. Bojesen SE. Telomeres and human health. J Intern Med. 2013;274(5):399-413.

16. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. Human molecular genetics. 2014;23(R1):R89-98.

17. Levy D, Neuhausen SL, Hunt SC, Kimura M, Hwang SJ, Chen W, et al. Genome-wide association identifies OBFC1 as a locus involved in human leukocyte telomere biology. Proc Natl Acad Sci U S A. 2010;107(20):9293-8.

18. Prescott J, Kraft P, Chasman DI, Savage SA, Mirabello L, Berndt SI, et al. Genome-wide association study of relative telomere length. PLoS One. 2011;6(5):e19635.

19. Mangino M, Hwang SJ, Spector TD, Hunt SC, Kimura M, Fitzpatrick AL, et al. Genome-wide meta-analysis points to CTC1 and ZNF676 as genes regulating telomere homeostasis in humans. Human molecular genetics. 2012;21(24):5385-94.

20. Pooley KA, Bojesen SE, Weischer M, Nielsen SF, Thompson D, Amin Al Olama A, et al. A genome-wide association scan (GWAS) for mean telomere length within the COGS project: identified loci show little association with hormone-related cancer risk. Human molecular genetics. 2013;22(24):5056-64.

21. Codd V, Nelson CP, Albrecht E, Mangino M, Deelen J, Buxton JL, et al. Identification of seven loci affecting mean telomere length and their association with disease. Nat Genet. 2013;45(4):422-7, 7e1-2.

22. Zhang C, Doherty JA, Burgess S, Hung RJ, Lindstrom S, Kraft P, et al. Genetic determinants of telomere length and risk of common cancers: a Mendelian randomization study. Human molecular genetics. 2015;24(18):5356-66.

23. Rode L, Nordestgaard BG, Bojesen SE. Long telomeres and cancer risk among 95 568 individuals from the general population. International journal of epidemiology. 2016;45(5):1634-43.

24. The Telomeres Mendelian Randomization Collaboration. Association Between Telomere Length and Risk of Cancer and Non-Neoplastic Diseases: A Mendelian Randomization Study. JAMA Oncology. 2017;3(5):636-51.

25. Seow WJ, Cawthon RM, Purdue MP, Hu W, Gao YT, Huang WY, et al. Telomere length in white blood cell DNA and lung cancer: a pooled analysis of three prospective cohorts. Cancer research. 2014;74(15):4090-8.

26. Lan Q, Cawthon R, Gao Y, Hu W, Hosgood HD, 3rd, Barone-Adesi F, et al. Longer telomere length in peripheral white blood cells is associated with risk of lung cancer and the rs2736100 (CLPTM1L-TERT) polymorphism in a prospective cohort study among women in China. PLoS One. 2013;8(3):e59230.

27. Shen M, Cawthon R, Rothman N, Weinstein SJ, Virtamo J, Hosgood HD, 3rd, et al. A prospective study of telomere length measured by monochrome multiplex quantitative PCR and risk of lung cancer. Lung cancer. 2011;73(2):133-7.

28. Sanchez-Espiridion B, Chen M, Chang JY, Lu C, Chang DW, Roth JA, et al. Telomere length in peripheral blood leukocytes and lung cancer risk: a large case-control study in Caucasians. Cancer research. 2014;74(9):2476-86.

29. Jang JS, Choi YY, Lee WK, Choi JE, Cha SI, Kim YJ, et al. Telomere length and the risk of lung cancer. Cancer science. 2008;99(7):1385-9.

30. Sun B, Wang Y, Kota K, Shi Y, Motlak S, Makambi K, et al. Telomere length variation: A potential new telomere biomarker for lung cancer risk. Lung cancer. 2015;88(3):297-303.

31. Zhang Y, Sturgis EM, Dahlstrom KR, Wen J, Liu H, Wei Q, et al. Telomere length in peripheral blood lymphocytes contributes to the development of HPV-associated oropharyngeal carcinoma. Cancer research. 2013;73(19):5996-6003.

32. Bau DT, Lippman SM, Xu E, Gong Y, Lee JJ, Wu X, et al. Short telomere lengths in peripheral blood leukocytes are associated with an increased risk of oral premalignant lesion and oral squamous cell carcinoma. Cancer. 2013;119(24):4277-83.

33. Wang Z, Zhu B, Zhang M, Parikh H, Jia J, Chung CC, et al. Imputation and subset-based association analysis across different cancer types identifies multiple independent risk loci in the TERT-CLPTM1L region on chromosome 5p15.33. Human molecular genetics. 2014.

34. Amos CI, Dennis J, Wang Z, Byun J, Schumacher FR, Gayther SA, et al. The OncoArray Consortium: a Network for Understanding the Genetic Architecture of Common Cancers. Cancer Epidemiol Biomarkers Prev. 2016.

35. Lesseur C, Diergaarde B, Olshan AF, Wunsch-Filho V, Ness AR, Liu G, et al. Genome-wide association analyses identify new susceptibility loci for oral cavity and pharyngeal cancer. Nat Genet. 2016;48(12):1544-50.

36. Li Y, Byun J, Cai G, Xiao X, Han Y, Cornelis O, et al. FastPop: a rapid principal component derived method to infer intercontinental ancestry using genetic data. BMC bioinformatics. 2016;17:122.

37. Kachuri L, Amos CI, McKay JD, Johansson M, Vineis P, Bueno-de-Mesquita HB, et al. Fine mapping of chromosome 5p15.33 based on a targeted deep sequencing and high density genotyping identifies novel lung cancer susceptibility loci. Carcinogenesis. 2016;37(1):96-105.

38. Weischer M, Bojesen SE, Cawthon RM, Freiberg JJ, Tybjaerg-Hansen A, Nordestgaard BG. Short telomere length, myocardial infarction, ischemic heart disease, and early death. Arterioscler Thromb Vasc Biol. 2012;32(3):822-9.

39. Burgess S, Thompson SG. Use of allele scores as instrumental variables for Mendelian randomization. International journal of epidemiology. 2013;42(4):1134-44.

40. Pierce BL, Ahsan H, Vanderweele TJ. Power and instrument strength requirements for Mendelian randomization studies using multiple genetic variants. International journal of epidemiology. 2011;40(3):740-52.

41. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. Genetic epidemiology. 2013;37(7):658-65.

42. Thompson JR, Minelli C, Abrams KR, Tobin MD, Riley RD. Meta-analysis of genetic studies using Mendelian randomization--a multivariate approach. Statistics in medicine. 2005;24(14):2241-54.

43. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. Genetic epidemiology. 2016;40(4):304-14.

44. VanderWeele TJ. A three-way decomposition of a total effect into direct, indirect, and interactive effects. Epidemiology. 2013;24(2):224-32.

45. Valeri L, Vanderweele TJ. Mediation analysis allowing for exposure-mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros. Psychol Methods. 2013;18(2):137-50.

46. VanderWeele TJ. Explanation in Causal Inference: Methods for Mediation and Interaction. New York, NY: Oxford University Press; 2015.

47. Bojesen SE, Pooley KA, Johnatty SE, Beesley J, Michailidou K, Tyrer JP, et al. Multiple independent variants at the TERT locus are associated with telomere length and risks of breast and ovarian cancer. Nat Genet. 2013;45(4):371-84, 84e1-2.

48. Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. International journal of epidemiology. 2017;46(6):1985-98.

49. McKay JD, Hung RJ, Gaborieau V, Boffetta P, Chabrier A, Byrnes G, et al. Lung cancer susceptibility locus at 5p15.33. Nat Genet. 2008;40(12):1404-6.

50. Pande M, Spitz MR, Wu X, Gorlov IP, Chen WV, Amos CI. Novel genetic variants in the chromosome 5p15.33 region associate with lung cancer risk. Carcinogenesis. 2011;32(10):1493-9.

51. McKay JD, Hung RJ, Han Y, Zong X, Carreras-Torres R, Christiani DC, et al. Large-scale association analysis identifies new lung cancer susceptibility loci and heterogeneity in genetic susceptibility across histological subtypes. Nat Genet. 2017;49(7):1126-32.

52. Brennan P, Hainaut P, Boffetta P. Genetics of lung-cancer susceptibility. The lancet oncology. 2011;12(4):399-408.

53. Samet JM, Avila-Tang E, Boffetta P, Hannan LM, Olivo-Marston S, Thun MJ, et al. Lung cancer in never smokers: clinical epidemiology and environmental risk factors. Clinical cancer research : an official journal of the American Association for Cancer Research. 2009;15(18):5626-45.

54. Couraud S, Zalcman G, Milleron B, Morin F, Souquet PJ. Lung cancer in never smokers--a review. European journal of cancer. 2012;48(9):1299-311.

55. Rafnar T, Sulem P, Stacey SN, Geller F, Gudmundsson J, Sigurdsson A, et al. Sequence variants at the TERT-CLPTM1L locus associate with many cancer types. Nat Genet. 2009;41(2):221-7.

56. Liu Z, Ma H, Wei S, Li G, Sturgis EM, Wei Q. Telomere length and TERT functional polymorphisms are not associated with risk of squamous cell carcinoma of the head and neck. Cancer Epidemiol Biomarkers Prev. 2011;20(12):2642-5.

57. Gu Y, Yu C, Miao L, Wang L, Xu C, Xue W, et al. Telomere length, genetic variants and risk of squamous cell carcinoma of the head and neck in Southeast Chinese. Scientific reports. 2016;6:20675.

58. Shay JW, Bacchetti S. A survey of telomerase activity in human cancer. European journal of cancer. 1997;33(5):787-91.

59. Bull CF, Mayrhofer G, O'Callaghan NJ, Au AY, Pickett HA, Low GK, et al. Folate deficiency induces dysfunctional long and short telomeres; both states are associated with hypomethylation and DNA damage in human WIL2-NS cells. Cancer prevention research. 2014;7(1):128-38.

60. Borah S, Xi L, Zaug AJ, Powell NM, Dancik GM, Cohen SB, et al. Cancer. TERT promoter mutations and telomerase reactivation in urothelial cancer. Science. 2015;347(6225):1006-10.

61. Zheng YL, Zhang F, Sun B, Du J, Sun C, Yuan J, et al. Telomerase enzymatic component hTERT shortens long telomeres in human cells. Cell Cycle. 2014;13(11):1765-76.

62. Martinez P, Thanasoula M, Munoz P, Liao C, Tejera A, McNees C, et al. Increased telomere fragility and fusions resulting from TRF1 deficiency lead to degenerative pathologies and increased cancer in mice. Genes & development. 2009;23(17):2060-75.

63. Rivera T, Haggblom C, Cosconati S, Karlseder J. A balance between elongation and trimming regulates telomere stability in stem cells. Nat Struct Mol Biol. 2017;24(1):30-9.

64. Pai SI, Westra WH. Molecular pathology of head and neck cancer: implications for diagnosis, prognosis, and treatment. Annu Rev Pathol. 2009;4:49-70.

65. von Zglinicki T. Role of oxidative stress in telomere length regulation and replicative senescence. Ann N Y Acad Sci. 2000;908:99-110.

66. Hohensinner PJ, Goronzy JJ, Weyand CM. Telomere dysfunction, autoimmunity and aging. Aging Dis. 2011;2(6):524-37.

67. Gillison ML, Chaturvedi AK, Anderson WF, Fakhry C. Epidemiology of Human Papillomavirus-Positive Head and Neck Squamous Cell Carcinoma. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2015;33(29):3235-42.

68. de Martel C, Plummer M, Vignat J, Franceschi S. Worldwide burden of cancer attributable to HPV by site, country and HPV type. International journal of cancer Journal international du cancer. 2017.

69. Wu YH, Graff RE, Passarelli MN, Hoffman JD, Ziv E, Hoffmann TJ, et al. Identification of Pleiotropic Cancer Susceptibility Variants from Genome-Wide Association Studies Reveals Functional Characteristics. Cancer Epidemiol Biomarkers Prev. 2018;27(1):75-85.

70. Gudmundsson J, Besenbacher S, Sulem P, Gudbjartsson DF, Olafsson I, Arinbjarnarson S, et al. Genetic correction of PSA values using sequence variants associated with PSA levels. Sci Transl Med. 2010;2(62):62ra92.

71. Kamatani Y, Matsuda K, Okada Y, Kubo M, Hosono N, Daigo Y, et al. Genome-wide association study of hematological and biochemical traits in a Japanese population. Nat Genet. 2010;42(3):210-5.

72. Fingerlin TE, Murphy E, Zhang W, Peljto AL, Brown KK, Steele MP, et al. Genome-wide association study identifies multiple susceptibility loci for pulmonary fibrosis. Nat Genet. 2013;45(6):613-20.

73. Low KC, Tergaonkar V. Telomerase: central regulator of all of the hallmarks of cancer. Trends Biochem Sci. 2013;38(9):426-34.

74. Swanson SA, Tiemeier H, Ikram MA, Hernan MA. Nature as a Trialist?: Deconstructing the Analogy Between Mendelian Randomization and Randomized Trials. Epidemiology. 2017;28(5):653-9.

75. VanderWeele TJ, Tchetgen Tchetgen EJ, Cornelis M, Kraft P. Methodological challenges in mendelian randomization. Epidemiology. 2014;25(3):427-35.

76. Friedrich U, Griese E, Schwab M, Fritz P, Thon K, Klotz U. Telomere length in different tissues of elderly patients. Mech Ageing Dev. 2000;119(3):89-99.

77. Saferali A, Lee J, Sin DD, Rouhani FN, Brantly ML, Sandford AJ. Longer telomere length in COPD patients with alpha1-antitrypsin deficiency independent of lung function. PLoS One. 2014;9(4):e95600.

78. Daniali L, Benetos A, Susser E, Kark JD, Labat C, Kimura M, et al. Telomeres shorten at equivalent rates in somatic tissues of adults. Nat Commun. 2013;4:1597.

79. Haycock PC, Burgess S, Wade KH, Bowden J, Relton C, Davey Smith G. Best (but oft-forgotten) practices: the design, analysis, and interpretation of Mendelian randomization studies. The American journal of clinical nutrition. 2016;103(4):965-78.

80. Hartwig FP, Borges MC, Horta BL, Bowden J, Davey Smith G. Inflammatory Biomarkers and Risk of Schizophrenia: A 2-Sample Mendelian Randomization Study. JAMA psychiatry. 2017;74(12):1226-33.

81. Carreras-Torres R, Johansson M, Haycock PC, Wade KH, Relton CL, Martin RM, et al. Obesity, metabolic factors and risk of different histological types of lung cancer: A Mendelian randomization study. PLoS One. 2017;12(6):e0177875.

82. Dimitrakopoulou VI, Tsilidis KK, Haycock PC, Dimou NL, Al-Dabhani K, Martin RM, et al. Circulating vitamin D concentration and risk of seven cancers: Mendelian randomisation study. Bmj. 2017;359:j4761.

83. Richmond RC, Hemani G, Tilling K, Davey Smith G, Relton CL. Challenges and novel approaches for investigating molecular mediation. Human molecular genetics. 2016;25(R2):R149-R56.

Table 1: Characteristics of the Toronto (MSH-PMH) and Copenhagen (CGPS) OncoArray studies that comprise the dataset for the development of genetic instruments for telomere length in chromosome 5p15.33

Characteristic and description		Toronto (MSH-PMH)		Copenhagen (CGPS)		Total	
		N	(%)	N	(%)	N	(%)
Age (years)	<50	135	(17.4)	287	(24.5)	422	(20.6
	50 to 59	241	(28.6)	259	(22.1)	500	(24.4)
	60 to 69	313	(35.0)	264	(22.5)	577	(28.1)
	70 to 79	143	(14.7)	237	(20.2)	380	(18.5)
	≥80	47	(4.3)	125	(10.7)	172	(8.4)
	Mean (SD)	61.0	(11.7)	61.3	(12.8)	61.2	(12.3)
Sex	Males	436	(49.6)	470	(40.1)	906	(44.2)
	Females	443	(50.4)	702	(59.9)	1145	(55.8)
Smoking status	Never smokers	438	(50.1)	410	(36.4)	848	(41.3)
	Ever smokers	436	(49.6)	717	(61.2)	1153	(56.2)
	Former smokers	366	(41.7)	717	(61.2)	1083	(52.8)
	Current smokers	59	(6.7)	0	(0)	59	(2.9)
	Unknown	5	(0.6)	45	(3.8)	50	(2.4)
	Mean cigarette pack-years (SD)	8.7	(17.2)	14.4	(20.2)	12.0	(19.2)
Total		879		1172		2051	(100.0)

Abbreviations:

CGPS	Copenhagen General Population Study
MSH-PMH	Mount Sinai Hospital-Princess Margaret Hospital study
SD	Standard deviation

Table 2: Genetic variants included in the novel 5p15.33 instrumental variable and their associations with the telomere length Z-score in the combined dataset (n=2051)

Variant	Gene	Alleles		EAF	Per-allele estimate		P-value
		Long TL	Other		β[a,b]	(SE)	
rs956942	*LINC01511*	A	G	2.4×10^{-3}	1.11	(0.29)	1.7×10^{-4}
Chr5:1383486	*CLPTM1L-SLC6A3*	A	G	4.9×10^{-4}	2.09	(0.65)	1.4×10^{-3}
Chr5:1404329	*SLC6A3*	T	C	9.8×10^{-4}	1.28	(0.46)	5.8×10^{-3}
Chr5:1501109	*LPCAT1*	A	G	7.4×10^{-4}	1.46	(0.53)	6.1×10^{-3}
Chr5:1297379	*TERT*	A	C/G	1.5×10^{-3}	0.68	(0.27)	0.01
rs80022192	*LINC01511*	G	A	4.9×10^{-4}	1.60	(0.65)	0.01
rs35033501	*TERT*	A	G	0.03	0.22	(0.09)	0.01
rs28363089	*SLC6A3*	A	G	0.03	0.23	(0.02)	0.01
Chr5:1434327	*SLC6A3*	A	T	0.99	0.89	(0.38)	0.02
Chr5:1402812	*SLC6A3*	T	C	4.9×10^{-4}	1.49	(0.65)	0.02
rs79717857	*CLPTM1L*	A	C	0.02	0.21	(0.09)	0.02
rs35334674	*TERT*	G	A	0.97	0.19	(0.08)	0.02
rs7733853	*LPCAT1*	A	G	0.24	0.08	(0.03)	0.02
rs72715516	*SLC6A3*	G	A	0.96	0.21	(0.10)	0.04

Abbreviations:

EAF	Effect allele frequency, where the effect allele is the long telomere allele
SE	Standard error
LINC01511	Long intergenic non-protein coding RNA 1511
CLPTM1L	Cleft lip and palate associated transmembrane protein 1-like
SLC6A3	Solute carrier family 6 member 3
LPCAT1	Lysophosphatidylcholine acyltransferase 1
TERT	Telomerase reverse transcriptase

[a] Linear regression models adjusted for age, sex, study, and ethnicity principal components

[b] Regression coefficients are standardized and correspond to a 1 standard deviation (1 unit) change in the telomere length Z-score, approximately 1000 base pairs

Table 3: Per-allele associations for the 5p15.33 genetic instrument and relevant telomere and cancer endpoints

Outcome	Sample size (Cases, Controls)		β[a] / OR[b]	(SE) / 95% CI	P-value	F statistic	R^2 (%)
Telomere Length		2051	0.14	(0.02)	2.6×10^{-9}	35.83	1.49
Telomere length in never smokers		848	0.18	(0.04)	7.0×10^{-6}	20.81	2.02
Smoking status (ever/never)		2051	-0.08	(0.06)	0.19	-	-
Cigarette pack-years		1101	0.40	(0.73)	0.59	0.29	0.00
Lung cancer	16396	13013	1.04	1.01, 1.07	4.89×10^{-3}	-	-
Adenocarcinoma	5690	13013	1.06	1.03, 1.10	1.4×10^{-3}	-	-
Squamous cell carcinoma	4045	13013	1.03	0.98, 1.07	0.23	-	-
Head and neck cancer	4415	5013	0.95	0.90, 1.00	0.04	-	-
Oral cavity	2284	5013	0.93	0.87, 0.98	0.01	-	-
Oropharynx	1849	5013	0.96	0.90, 1.03	0.26	-	-
Never smokers							
Lung cancer	1619	3923	1.06	0.99, 1.14	0.08	-	-
Adenocarcinoma	836	3923	1.12	1.02, 1.22	0.02	-	-
Head and neck cancer	773	1827	0.85	0.77, 0.95	3.8×10^{-3}	-	-
Alcohol non-drinkers							
Head and neck cancer	614	795	0.86	0.74, 0.99	0.04	-	-

Abbreviations:

R^2 Coefficient of determination estimating the proportion of the variance in the telomere length Z-score that is explained by the 5p15.33 genetic instrument

SE Standard error

TL Telomere length

[a] Linear regression models were adjusted for age, sex, study, and top 5 ethnicity principal components

[b] Logistic regression models were adjusted for age, sex, study, and top 10 ethnicity principal components

Table 4: Mendelian Randomization estimates of the causal odds ratios for lung and head and neck cancers per 1000 base pair increase in telomere length

Outcome	Cases	Controls	Estimation Method								
			Maximum Likelihood			Inverse Variance Weighted			Weighted Median Estimator		
			OR[a]	95% CI	P-value	OR[a]	95% CI	P-value	OR[a]	95% CI	P-value
Lung cancer	16396	13013	1.41	1.20, 1.65	2.0×10^{-5}	1.39	1.21, 1.60	3.7×10^{-6}	1.37	1.12, 1.67	2.0×10^{-3}
Adenocarcinoma	5690	13013	1.92	1.51, 2.45	1.3×10^{-7}	1.83	1.51, 2.22	5.5×10^{-10}	1.63	1.23, 2.16	6.5×10^{-4}
Squamous	4045	13013	1.04	0.83, 1.29	0.74	1.04	0.83, 1.29	0.74	1.09	0.82, 1.46	0.57
Small cell	1846	13013	1.03	0.76, 1.39	0.86	1.03	0.76, 1.38	0.86	0.96	0.66, 1.38	0.82
Head and neck cancer	4415	5013	0.90	0.70, 1.15	0.39	0.90	0.70, 1.15	0.41	0.71	0.51, 0.98	0.04
Oral cavity	2284	5013	0.88	0.65, 1.19	0.40	0.88	0.65, 1.19	0.40	0.67	0.44, 1.03	0.07
Oropharynx	1849	5013	0.83	0.59, 1.16	0.28	0.83	0.60, 1.16	0.28	0.72	0.46, 1.12	0.14
Ever smokers											
Lung cancer	14498	8815	1.36	1.14, 1.63	5.3×10^{-4}	1.36	1.15, 1.60	2.6×10^{-4}	1.31	1.05, 1.63	0.02
Adenocarcinoma	4754	8815	1.72	1.33, 2.24	4.2×10^{-5}	1.66	1.33, 2.07	5.2×10^{-6}	1.71	1.26, 2.32	6.1×10^{-4}
Squamous	3835	8815	1.06	0.84, 1.35	0.60	1.06	0.84, 1.35	0.61	1.08	0.80, 1.47	0.63
Head and neck	3108	2865	1.12	0.79, 1.58	0.54	1.11	0.79, 1.56	0.54	0.91	0.60, 1.39	0.69
Never smokers											
Lung cancer	1619	3923	1.78	1.22, 2.61	3.1×10^{-3}	1.76	1.23, 2.52	2.0×10^{-3}	1.55	0.98, 2.46	0.06
Adenocarcinoma	836	3923	2.68	1.70, 4.24	2.4×10^{-5}	2.68	1.70, 4.24	2.4×10^{-5}	2.24	1.18, 4.27	0.01
Squamous	149	3923	0.72	0.26, 1.97	0.52	0.72	0.26, 1.95	0.51	0.80	0.22, 2.90	0.75
Head and neck	773	1827	0.72	0.42, 1.22	0.22	0.72	0.42, 1.22	0.22	0.71	0.32, 1.55	0.39
Early onset (≤50 years)											
Lung cancer	1868	1557	1.68	1.07, 2.62	0.02	1.67	1.08, 2.59	0.02	1.76	0.98, 3.22	0.06
Alcohol non-drinkers											
Head and neck	614	795	0.76	0.37, 1.56	0.45	0.76	0.37, 1.57	0.46	0.45	0.17, 1.16	0.10

Abbreviations:

CI Confidence Intervals

OR Odds ratio

[a] Regression models for each genetic instrument were adjusted for age, sex, study, and the top 10 ethnicity principal components

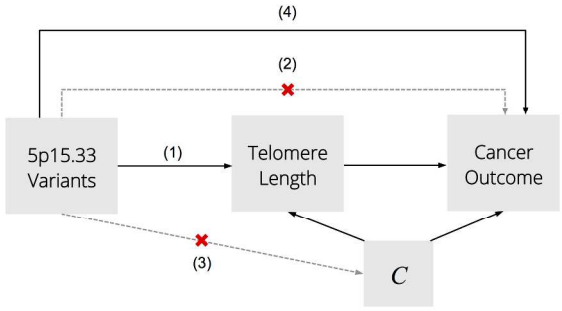

Figure 1: Conceptual diagram of Mendelian Randomization and mediation analyses. Mendelian Randomization is based the following assumptions (1–3): the genetic variant is strongly associated with telomere length; there is no direct association between the instrument and cancer outcome, except through telomere length; the genetic instrument is independent of any confounders (C). Mediation analyses of the 5p15.33 instrument for telomere length and 5p15.33 susceptibility variants test for the presence of direct effects (4), and quantify how much of the total genetic effect on lung cancer risk is mediated by telomere length

254x190mm (300 x 300 DPI)

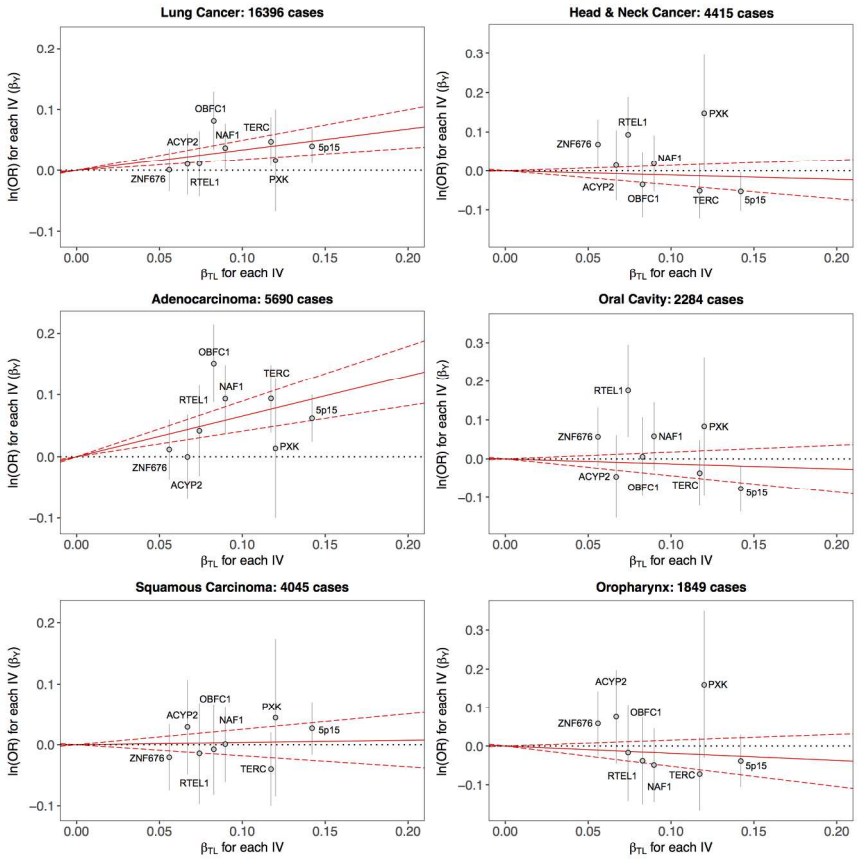

Figure 2: Scatter plots showing the association estimates for telomere length (β_{TL}) and cancer risk (β_Y) for each instrumental variable (IV), overlaid on the causal log odds ratio for the effect of increasing telomere length on cancer risk (solid red line) and corresponding 95% confidence intervals (dotted red lines), estimated using the likelihood-based method

190x254mm (300 x 300 DPI)

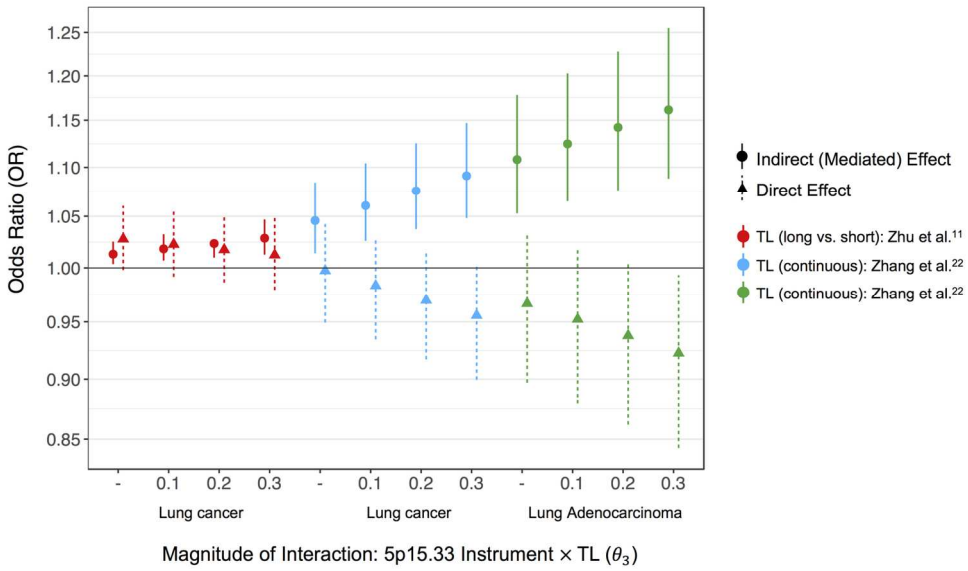

Figure 3: Odds ratio (OR) plot summarizing the direct effects (triangle, dotted line) and indirect effects (circle, solid line) of the 5p15.33 genetic instrument on lung cancer risk. Estimates of the direct and indirect effects are presented across different levels of interaction and for different versions of the mediator (dichotomous and continuous), indicated by different colours.

149x91mm (300 x 300 DPI)