THE UNIVERSITY OF CHICAGO

A STUDY OF TELOMERE LENGTH AS A PROSPECTIVE BIOMARKER OF CANCER RISK

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES

AND THE PRITZKER SCHOOL OF MEDICINE

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF PUBLIC HEALTH SCIENCES

BY

CHENAN ZHANG

CHICAGO, ILLINOIS

JUNE 2016

Dedication:

To my parents, who raised their daughter with the firm belief that no goal is beyond reach,

To my past and present mentors, who provided me with the tools I needed to face this challenge,

And to my husband – how joyful I am to have you by my side as I reach for this goal!

CONTENTS

LIST OF TABLES

LIST OF SUPPLEMENTARY TABLES

LIST OF FIGURES

LIST OF SUPPLEMENTARY FIGURES

ACKNOWLEDGMENTS

This dissertation work became a reality with the support and help of many individuals. I would like to extend my sincerest thanks to all of them.

I want to thank my advisor, Dr. Brandon Pierce, for his exceptional mentorship. I couldn't have imagined the wealth of knowledge, depth of expertise, generous attention, and friendly rapport that he would share with me when I started a rotation with this young professor during my first year as a Ph.D. student. He has impressed on me the qualities of a good mentor and researcher that I will carry with me and emulate as I embark on my own journey in academia. It has been an honor being his first Ph.D. student.

I also want to thank Dr. Diane Lauderdale, who expanded my training in the area of biodemography as the chair of my dissertation committee, and advisor of my National Institute on Aging training fellowship. Dr. Lauderdale's knack for pointed questions and elegant suggestions has cultivated my own discerning eye towards issues in both epidemiology and biodemography.

I am grateful to committee members Dr. Habibul Ahsan, for his valuable insights on the Bangladesh cohort study, Dr. Lin Chen for being the first to teach me what a regression is, and Dr. Kenan Onel, for encouraging me to become acquainted with the biology underlying the epidemiologic question being investigated.

Thank you to Dr. Muhammad Kibriya, Dr. Farzana Jasmine, and the rest of the Ahsan Lab for producing the valuable data for me to use. Thank you to Maria Argos and Lin Tong for patiently bringing me up to date on the Bangladesh cohort study data structure.

Thank you also to Dr. Benjamin Lahey, my academic advisor during the first two years, Michele Thompson, for her constant support and sympathetic ear, Kimberly Cox and Kelly

Marcias for successfully wrangling meeting times among very busy professors, and Dr. Tamara Konetzka and Dr. Lianne Kurina for their encouragement and availability – I always felt I was in good hands as a student in the department.

Thank you to Dr. Linda Waite, Dr. Melissa Howe, and the rest of the Committee on Demographic Training for the generous pre-doctoral fellowship support in the Demography and Economics of Aging and the engaging weekly seminars and post-mortem discussions.

Molly Scannell, I couldn't have hoped for a better cohort-mate, class-mate, office-mate, and friend than you. I will always look back at our lively office discussions with fondness. To classmates in the department of Public Health Sciences past and present, thank you for your companionship, commiseration, and support in the last five years. To friends Jaclyn Wong, Kevin Muite, and Carolina Mora, thank you for being terrific balancing forces in my life!

To Loe Larsson and David Beine, thank you for being the family we needed during our years in Chicago together. To my dear Gustav Larsson, thank you for always staying up with me when I had to work late. To my parents Dr. Mingyuan Zhang and Dr. Xiaoping Li, thank you for always offering to proofread my drafts even when you no longer understood everything I wrote – and still catching things that I missed!

I am grateful to have so much love and support in my life!

ABSTRACT

This dissertation is comprised of two peer-reviewed articles and one manuscript, aiming to identify knowledge gaps in terms of the role of telomere length as a biomarker of cancer.

Individuals with relatively short telomeres in peripheral blood cells have been shown in epidemiological studies to be at increased risk for mortality and cancer, leading investigators to hypothesize that telomere length and attrition over the life-course is a critical mechanism underlying cancer and many other age-related health conditions. However, such associations are not consistent across all cancers or even within cancer types, with some studies showing null, or even opposite associations. Furthermore, due to the cross-sectional nature of case-control studies, from which most associations of telomere length and cancer risk are obtained, it is possible that telomere shortening occurs after diagnosis (due to the cancer itself or treatment effects), or is confounded by other common risk factors, and therefore may not be a true biomarker of subsequent cancer risk.

I employed three strategies for addressing these potential issues: 1) use genetic polymorphisms that influence telomere length as proxies for telomere length to estimate associations between telomere length and cancer risk; 2) perform careful adjustment of potential confounders of the telomere length-cancer association, including the assessment of time-varying effects on smoking, a potential telomere length correlate; and 3) utilize a prospective study design in which baseline telomere length is measured prior to the development of any adverse health outcomes.

In the first study of this dissertation (Chapter 2), I used a Mendelian randomization approach to estimate the associations between nine telomere length-associated SNPs and risk for five common cancer types (breast, lung, colorectal, ovarian and prostate cancer, including

subtypes) using data on 51,725 cases and 62,035 controls. The long TL genetic score was significantly associated with increased risk of lung adenocarcinoma (P=6.3x10^{-15}), with an estimated odds ratio of 2.78. Under Mendelian randomization assumptions, this estimate was interpreted as the odds ratio for lung adenocarcinoma corresponding to a 1000 base pair increase in telomere length. The SNP score was not associated with other cancer types or subtypes.

In the second study of this dissertation (Chapter 3), I estimated the associations between telomere length (measured by quantitative PCR using saliva DNA) and concurrent and past smoking status, reported biennially for up to 16 years prior to telomere length measurement, using data from the Health and Retirement Study (n=5,624). Smoking was associated with shorter telomere length when using prospective data on smoking status among men and women, but the association was strongly attenuated in cross-sectional analyses among men. This attenuation was largely due to a higher rate of smoking cessation during the study period among males with shorter telomere length compared to males with longer telomere length. Short telomere length was also associated with poorer overall health in men, suggesting that male smokers with short telomere length were more likely to quit smoking due to poor health. Analyses of years since cessation, smoking duration, and pack-years all support the hypothesis that increased cigarette use shortens telomere length.

In the third study of this dissertation (Chapter 4), I investigated the association between arsenic exposure and telomere length, and the association between telomere length and incident arsenical skin lesion in the Health Effects of Arsenic Longitudinal Study (HEALS) in Araihazar, Bangladesh. In this prospective study, baseline telomere length was assessed in skin lesion-free subjects who were followed for up to 12 years. No association was observed between baseline arsenic exposure and telomere length. However, we observed higher incident skin lesion risk

with shorter telomere length (P_{trend}=4.6x10^{-5}), with odds ratios of 3.05, 1.30, and 1.21 for the first (shortest), second, and third telomere length quartiles compared to the longest.

The role of telomere length as a biomarker of cancer appears to be striking yet nuanced. Addressing the existing knowledge gaps is a critical step towards clarifying the causal relationship between telomere length and cancer, and ultimately, improving cancer prediction and prevention.

CHAPTER 1: BACKGROUND AND INTRODUCTION

TELOMERE STRUCTURE AND FUNCTION

Telomeres are DNA-protein complexes at chromosome ends that help maintain genome stability by protecting DNA from damage and fusion. The DNA component is a 6-base repeat sequence (TTAGGG) ranging from 2-15 kb that shortens with each cell division due to a phenomenon known as the "end replication problem" (1). In differentiated cells, telomere shortening eventually leads to loss of telomere protection and genome instability, typically triggering cell senescence or programmed cell death (2). In stem and progenitor cells, the telomerase enzyme elongates telomeres, enabling prolonged survival (3). Telomerase is also activated in >90% of human tumors (4), which typically have short telomeres (a cause of genome instability) that are stabilized by telomerase (promoting proliferation and survival) (5,6).

In yeast, telomerase deletion causes rapid telomere shortening that results in increased mutation rates, chromosomal rearrangements, and deletions (7). In telomerase-deficient murine models, short telomeres lead to genomic instability, deficient DNA repair, increased sensitivity to ionizing radiation, and cancer (8–10). These models also suggest that the shortest telomere is more important than the average telomere length (TL) with respect to cell viability and chromosomal stability (8). In genomically unstable cells with short telomeres, activation of telomerase can serve as a transforming event that allows continued survival and potential tumor progression. Tumor cells have extremely short telomeres, yet they are stable, typically due to activation of telomerase, which is expressed at higher levels in cancerous compared to non-cancerous tissue (11–14).

Genome-wide association (GWA) studies of TL have identified several regions containing variants associated with TL (15–17), including the *TERT* (telomerase reverse transcriptase) region on 5p15.33. Furthermore, GWA studies of cancer risk have shown that variants residing near *TERT* influence risk for multiple cancer types, including lung (18,19), prostate (18,19), and colorectal (20) cancer. The critical role of telomeres and telomerase in carcinogenesis has led many investigators to hypothesize that short TL in peripheral blood (assumed to be a proxy for TL in cancer-prone tissue) is a risk factor for cancer (21).

DETERMINANTS OF TELOMERE LENGTH IN HUMANS

Average TL of peripheral blood cells differ between individuals due to inter-individual variation of TL inherited at conception (22–24), and intra-individual variation of TL attrition attributable to genetic and environmental factors (25–27). Previous studies have also described variation in TL within somatic tissues between proliferative (e.g. blood and skin) and minimally proliferative tissues (e.g. muscle and fat), presumably due to the greater TL attrition in more rapidly dividing proliferative tissues during early life (28–30).

Previous findings show strong evidence that age, sex, and race are associated with TL (31), and that paternal age and genetic variants are determinants of TL (15,16,32–34). Women have longer TL than men, a phenomenon that has been hypothesized to be due to estrogen's reduction of oxidative stress (35) and stimulation of telomere lengthening enzyme telomerase (36). African Americans have been observed to have longer TL in prior studies, speculated to be in part due to fewer replications of hematopoietic stem cells and progenitor cells in subjects of African ancestry (37,38), but also potentially due to polygenetic adaptation (39). Prior studies on

the association of older paternal age and longer TL have attributed the phenomenon to the continuous increase of TL during the development of male germ cells throughout a man's life course (40).

Individuals of low socioeconomic (SES) status are at higher risk for mortality and diseases of aging, including coronary heart disease and some cancers (41). Previous studies have primarily observed that low SES is associated with shorter TL (38,42,43), hypothesized to be due to factors such as chronic financial strain, hazardous work conditions or adverse home environments. However, null associations (44,45) have also been observed, with one study reporting an association between TL and education attainment, but not current income.

Measures of obesity (including body mass index (BMI), weight, weight change, waist circumference, hip circumference, percent body fat, visceral fat, subcutaneous fat, adiponectin, leptin, and genetic determinants of BMI) were either inversely associated (38,46–56) or not associated (50,57,58) with TL in prior studies, although a positive association was also reported in the Kaiser Permanente Genetic Epidemiology Research Study on Adult Health and Aging, a multi-ethnic cohort of >100,000 participants (59).

Smoking is a strong risk factor for overall mortality and many adverse health outcomes (60). Multiple prior studies have reported an association between smoking and short TL (61,62,46,48,63–68), although numerous also report no association (69–71,47,44,72). Heavy alcohol use has been associated with higher mortality and adverse health outcomes (73), but the opposite has been observed for moderate consumption, with an overall J-shaped relationship attributed to a combination of adverse and beneficial effects (74,75). Previous studies have reported either no association between alcohol consumption and TL (44,70), or a decrease in TL with prospectively-measured alcohol consumption (76). However, a previous study of TL in

yeast found that ethanol exposure increased TL (77), supporting the hypothesis that the nature of the effects of exposures on TL depends on the biological mechanisms of the stressor.

Many health benefits are attributed to physical exercise, even independent of its correlation with lower BMI (78). Physical exercise has been shown to up-regulate anti-inflammatory processes (79), decreasing the risk of some oxidative stress-related diseases, which may explain its association with longer TL in prior studies (80,81).

TELOMERE LENGTH AS A BIOMARKER FOR CANCER: THE KNOWLEDGE GAPS

Many epidemiological studies have reported associations between short relative TL measured in peripheral blood cells, and increased risk for various cancer types, including lung (82–85), ovarian (86,87), colorectal (88), and breast cancers (89–91), leading investigators to hypothesize that TL and TL attrition over the life-course is a critical mechanism underlying cancer and many other age-related health conditions. However, such associations are not consistent across all cancers or even within cancer types, with some studies showing null, U-shaped, or inverse associations (90,92–98). Furthermore, due to the retrospective nature of case-control studies from which many of these association estimates are obtained, it is unclear whether telomere shortening occurs after diagnosis, potentially due to cancer progression or treatment (94,96,97). One striking example of the different conclusions drawn from retrospective versus prospective studies of TL and cancer risk is a meta-analysis by Wentzensen et al. (99), which noted the lack of association between TL and cancer risk in prospective studies compared to an association between short TL and increased cancer risk in retrospective studies. Additional

studies addressing potential issues in reverse causality and residual confounding need to be performed to further clarify the role of TL as a biomarker for cancer.

Another gap in the literature on the role of TL as a biomarker of cancer is the lack of systematic evaluation of the correlates of TL, particularly among older individuals. Although many studies of TL and disease risk adjust for age and sex, most do not comprehensively adjust for other potential confounders such as sociodemographic or lifestyle covariates. While a recent review paper summarizes TL-associated phenotypes from different studies (31), few have comprehensively examined TL correlates in a single population as was done in the Copenhagen General Population Study (100). Furthermore, some correlates of TL, such as socioeconomic status or lifestyle factors may be time-varying in nature, resulting in potential confounding in cross-sectional analyses of correlates with TL. Efforts to comprehensively characterize TL correlates by simultaneous adjustment of covariates across multiple domains (including demographic, socioeconomic and lifestyle) with attention to time-varying covariates are necessary because inadequate adjustments for correlates of TL may confound associations between TL and cancer risk, resulting in biased associations. Because studies of TL and age-related diseases are typically conducted among older individuals, a more complete understanding of TL correlates in an older population would allow for better confounder identification and adjustment.

Also lacking from the literature on TL and cancer is data from non-Western populations. Although there is a sizable body of epidemiologic literature investigating the association between TL and cancer risk, these studies have primarily been conducted in populations of European (17,84,86,87,94–97) and East Asian descent (82,83,90,92,98). Inclusion of more diverse

populations are necessary to more fully understand TL determinants and associations in context of unique genetic, demographic, and lifestyle profiles.

STRATEGIES FOR ADDRESSING THE KNOWLEDGE GAPS

The primary gaps in knowledge were addressed in Chapters 2-4 of the dissertation. In Chapter 2, a Mendelian randomization approach was used to estimate the associations between nine TL-associated SNPs and risk for five common cancer types (breast, lung, colorectal, ovarian and prostate cancer, including subtypes) using data from Genetic Associations and Mechanisms in Oncology (GAME-ON), a large consortium of genome-wide association studies on 51,725 cases and 62,035 controls. Genetic variants are randomly assigned at conception, and therefore not susceptible to the effects of potentially confounding environmental factors or issues of reverse causality. By using genetic variants as proxies of TL, this method addresses issues that may bias retrospective studies specifically and observational studies in general.

In Chapter 3, a comprehensive approach was taken to simultaneously adjust for multiple demographic, socioeconomic, and lifestyle variables in association with TL in the Health and Retirement Study (HRS), a nationally representative longitudinal study of the health and economic well-being of aging individuals. Furthermore, the time-varying associations between smoking and TL were analyzed using smoking statuses reported every two years for up to 16 years prior to DNA collection for TL measurement. This comprehensive characterization of correlates of TL will allow for evidence-based confounder selection and adjustment for future

observational studies of TL and cancer risk, particularly in context of smoking, a major risk factor for cancer.

In Chapter 4, a prospective study design was employed in which baseline TL was assessed prior to any cancer development or other adverse health outcomes. Subjects in this study belonged to the Health Effects of Arsenic Longitudinal Study (HEALS) in Araihazar, Bangladesh, with individuals followed for up to 12 years for various outcomes, including skin lesions, nonmelanoma skin cancer, and mortality. In this chapter, TL association with arsenic exposure (as well as other demographic, socioeconomic, and lifestyle factors) were characterized in a Bangladesh cohort study. In addition, association between prospectively measured TL and subsequent skin lesion incidence was investigated in a nested case-control study. The prospective design of this study reduces potential bias due to reverse causality, and analysis of the Bangladeshi cohort provides a unique opportunity to investigate TL determinants and associations in a population of South Asian descent with a unique dietary profile and genetic background and a shorter life expectancy.

RATIONALE

Why examine non-genetic determinants of telomere length?

It is necessary to first establish that non-genetic determinants of telomere length observed in our studies are consistent with previous findings. This can serve to confirm the validity of TL measured in our laboratory and help to characterize sources of variation in our TL measure attributable to study design, covariates, or study-specific factors. Additionally, while the primary question is the effect of telomere length on cancer, it is also important to characterize the

possible determinants of TL as potential modifiable risk factors for cancer as a preliminary step for exploring subsequent cancer prevention measures.

Why use both genetic variants and molecular biomarkers as measures of telomere length?

Measures of telomere length through Southern blot analysis of the terminal restriction fragment length (101), and by RT-PCR (102) are well-validated methods. However, the analysis of telomere length association in epidemiologic observational studies, no matter how accurately measured, can be confounded by unmeasured covariates, and it is impossible to design a randomized controlled trial to test the effect of telomere length on health outcomes. Instead, genetic variants for telomere length can be compiled into a multi-SNP risk score (an "instrumental variable") used to indirectly to assess the causal relationship between TL and cancer while sidestepping the issues of confounding associated with analysis of observational data (103).

Why use multiple datasets?

The use of large case-control genome-wide association study data allows the analysis of large sample sizes of cancer cases, resulting in a better powered study to examine the relationship between TL-related SNPs and cancer risk. This is particularly beneficial in the instance of the post-genome wide association consortium study, GAME-ON, which has compiled summary data of multiple cancer sites from multiple groups, with up to tens of thousands of cases per analysis. The use of HRS, a nationally-representative cohort of older subjects to characterize correlates of TL is important for two reasons: 1) the need to characterize TL as a biomarker for cancer is particularly relevant to older individuals, who are more likely to experience aging-related

adverse health outcomes such as cancer; 2) older individuals represent an important age group to study given the longer period of time over which environmental effects on TL may accumulate. The use of the Bangladeshi prospective nested case-control design avoids or minimizes several potential biases of retrospective case-control studies (including survival bias and reverse causation). Furthermore, the additional use of multiple ethnic groups is valuable for investigating telomere length associations in groups of different genetic and lifestyle profiles, which may offer unique insights in heterogeneous effects across groups, or further strengthen the evidence of association for determinants that remain homogenous across groups.

SIGNIFICANCE

Should TL or its genetic proxies prove to be predictive of cancer risk, future cancer prevention efforts could consider TL as a novel risk measure, or potentially a pharmacological target for cancer prevention. Comparison of the genetic and environmental factors associated with TL in the Bangladeshi study and the two primarily Caucasian studies will improve our understanding of how TL determinants vary between ethnic populations, allowing the future development of targeted interventions. Addressing these knowledge gaps is a critical step towards clarifying the causal relationship between TL and cancer, with the long-term goal of using this knowledge to improve cancer prediction and prevention.

REFERENCES

1. Watson JD. Origin of concatemeric T7 DNA. Nature New Biol. 1972;239.

2. d'Adda di Fagagna F. Living on a break: cellular senescence as a DNA-damage response. Nat Rev Cancer. 2008;8:512–22.

3. Shawi M, Autexier C. Telomerase, senescence and ageing. Mech Ageing Dev. 2008;129:3–10.

4. Stewart SA, Weinberg RA. Telomeres: Cancer to Human Aging. Annu Rev Cell Dev Biol. 2006;22:531–57.

5. Blackburn EH, Greider CW, Szostak JW. Telomeres and telomerase: the path from maize, Tetrahymena and yeast to human cancer and aging. Nat Med. 2006;12:1133–8.

6. O'Sullivan RJ, Karlseder J. Telomeres: protecting chromosomes against genome instability. Nat Rev Mol Cell Biol. 2010;11:171–81.

7. Hackett JA, Feldser DM, Greider CW. Telomere Dysfunction Increases Mutation Rate and Genomic Instability. Cell. 2001;106:275–86.

8. Hemann MT, Strong MA, Hao L-Y, Greider CW. The Shortest Telomere, Not Average Telomere Length, Is Critical for Cell Viability and Chromosome Stability. Cell. 2001;107:67–77.

9. Artandi SE, Chang S, Lee S-L, Alson S, Gottlieb GJ, Chin L, et al. Telomere dysfunction promotes non-reciprocal translocations and epithelial cancers in mice. Nature. 2000;406:641–5.

10. Wong K-K, Chang S, Weiler SR, Ganesan S, Chaudhuri J, Zhu C, et al. Telomere dysfunction impairs DNA repair and enhances sensitivity to ionizing radiation. Nat Genet. 2000;26:85–8.

11. Shay JW, Bacchetti S. A survey of telomerase activity in human cancer. Spec Issue Telomeres Telomerase Cancer. 1997;33:787–91.

12. Kim N, Piatyszek M, Prowse K, Harley C, West M, Ho P, et al. Specific association of human telomerase activity with immortal cells and cancer. Science. 1994;266:2011–5.

13. Shammas MA, Qazi A, Batchu RB, Bertheau RC, Wong JYY, Rao MY, et al. Telomere Maintenance in Laser Capture Microdissection–Purified Barrett's Adenocarcinoma Cells and Effect of Telomerase Inhibition In vivo. Clin Cancer Res. 2008;14:4971–80.

14. Gertler R, Doll D, Maak M, Feith M, Rosenberg R. Telomere length and telomerase subunits as diagnostic and prognostic biomarkers in Barrett carcinoma. Cancer. 2008;112:2173–80.

15. 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;45:422–7.

16. Mangino M, Hwang S-J, 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. Hum Mol Genet. 2012;21:5385–94.

17. 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. Hum Mol Genet. 2013;22:5056–64.

18. 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;41:221–7.

19. Yin J, Li Y, Yin M, Sun J, Liu L, Qin Q, et al. TERT-CLPTM1L Polymorphism rs401681 Contributes to Cancers Risk: Evidence from a Meta-Analysis Based on 29 Publications. PLoS ONE. 2012;7:e50650.

20. Peters U, Hutter C, Hsu L, Schumacher F, Conti D, Carlson C, et al. Meta-analysis of new genome-wide association studies of colorectal cancer risk. Hum Genet. 2012;131:217–34.

21. Willeit P, Willeit J, Mayr A, et al. Telomere length and risk of incident cancer and cancer mortality. JAMA. 2010;304:69–75.

22. Graakjaer J, Pascoe L, Der-Sarkissian H, Thomas G, Kolvraa S, Christensen K, et al. The relative lengths of individual telomeres are defined in the zygote and strictly maintained during life. Aging Cell. 2004;3:97–102.

23. Graakjaer J, Der-Sarkissian H, Schmitz A, Bayer J, Thomas G, Kolvraa S, et al. Allele-specific relative telomere lengths are inherited. Hum Genet. 2006;119:344–50.

24. De Meyer T, Rietzschel ER, De Buyzere ML, De Bacquer D, Van Criekinge W, De Backer GG, et al. Paternal age at birth is an important determinant of offspring telomere length. Hum Mol Genet. 2007;16:3097–102.

25. Aviv A. The Epidemiology of Human Telomeres: Faults and Promises. J Gerontol A Biol Sci Med Sci. 2008;63:979–83.

26. Cassidy A, De Vivo I, Liu Y, Han J, Prescott J, Hunter DJ, et al. Associations between diet, lifestyle factors, and telomere length in women. Am J Clin Nutr. 2010;91:1273–80.

27. Slagboom P, Droog S, Boomsma D. Genetic determination of telomere size in humans: a twin study of three age groups. Am J Hum Genet. 1994;55:876–82.

28. Chang E, Harley CB. Telomere length and replicative aging in human vascular tissues. Proc Natl Acad Sci. 1995;92:11190–4.

29. Weng NP, Levine BL, June CH, Hodes RJ. Human naive and memory T lymphocytes differ in telomeric length and replicative potential. Proc Natl Acad Sci U S A. 1995;92:11091–4.

30. 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.

31. Sanders JL, Newman AB. Telomere Length in Epidemiology: A Biomarker of Aging, Age-Related Disease, Both, or Neither? Epidemiol Rev. 2013;35:112–31.

32. Broer L, Codd V, Nyholt DR, Deelen J, Mangino M, Willemsen G, et al. Meta-analysis of telomere length in 19,713 subjects reveals high heritability, stronger maternal inheritance and a paternal age effect. Eur J Hum Genet. 2013;21:1163–8.

33. Njajou OT, Cawthon RM, Damcott CM, Wu S-H, Ott S, Garant MJ, et al. Telomere length is paternally inherited and is associated with parental lifespan. Proc Natl Acad Sci. 2007;104:12135–9.

34. Kimura M, Cherkas LF, Kato BS, Demissie S, Hjelmborg JB, Brimacombe M, et al. Offspring's leukocyte telomere length, paternal age, and telomere elongation in sperm. PLoS Genet. 2008;4:e37.

35. Aviv A. Telomeres, sex, reactive oxygen species, and human cardiovascular aging. J Mol Med. 2002;80:689–95.

36. Kyo S, Takakura M, Kanaya T, Zhuo W, Fujimoto K, Nishio Y, et al. Estrogen Activates Telomerase. Cancer Res. 1999;59:5917–21.

37. Hunt SC, Chen W, Gardner JP, Kimura M, Srinivasan SR, Eckfeldt JH, et al. Leukocyte telomeres are longer in African Americans than in whites: the National Heart, Lung, and Blood Institute Family Heart Study and the Bogalusa Heart Study. Aging Cell. 2008;7:451–8.

38. Needham BL, Adler N, Gregorich S, Rehkopf D, Lin J, Blackburn EH, et al. Socioeconomic status, health behavior, and leukocyte telomere length in the National Health and Nutrition Examination Survey, 1999-2002. Soc Sci Med 1982. 2013;85:1–8.

39. Hansen MEB, Hunt SC, Stone RC, Horvath K, Herbig U, Ranciaro A, et al. Shorter Telomere Length in Europeans than in Africans due to Polygenetic Adaptation. Hum Mol Genet [Internet]. 2016; Available from: http://hmg.oxfordjournals.org/content/early/2016/03/01/hmg.ddw070.abstract

40. Aviv A, Susser E. Leukocyte Telomere Length and the Father's Age Enigma: Implications for Population Health and for Life Course. Int J Epidemiol. 2013;42:457–62.

41. Mackenbach JP, Stirbu I, Roskam A-JR, Schaap MM, Menvielle G, Leinsalu M, et al. Socioeconomic Inequalities in Health in 22 European Countries. N Engl J Med. 2008;358:2468–81.

42. Steptoe A, Hamer M, Butcher L, Lin J, Brydon L, Kivimäki M, et al. Educational attainment but not measures of current socioeconomic circumstances are associated with leukocyte telomere length in healthy older men and women. Brain Behav Immun. 2011;25:1292–8.

43. Cherkas LF, Aviv A, Valdes AM, Hunkin JL, Gardner JP, Surdulescu GL, et al. The effects of social status on biological aging as measured by white-blood-cell telomere length. Aging Cell. 2006;5:361–5.

44. Batty GD, Wang Y, Brouilette SW, Shiels P, Packard C, Moore J, et al. Socioeconomic status and telomere length: the West of Scotland Coronary Prevention Study. J Epidemiol Community Health. 2009;63:839–41.

45. Adams J, Martin-Ruiz C, Pearce MS, White M, Parker L, Von Zglinicki T. No association between socio-economic status and white blood cell telomere length. Aging Cell. 2007;6:125–8.

46. Valdes A, Andrew T, Gardner J, Kimura M, Oelsner E, Cherkas L, et al. Obesity, cigarette smoking, and telomere length in women. The Lancet. 20;366:662–4.

47. Nordfjäll K, Eliasson M, Stegmayr B, Melander O, Nilsson P, Roos G. Telomere Length Is Associated With Obesity Parameters but With a Gender Difference. Obesity. 2008;16:2682–9.

48. Parks CG, Miller DB, McCanlies EC, Cawthon RM, Andrew ME, DeRoo LA, et al. Telomere length, current perceived stress, and urinary stress hormones in women. Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol. 2009;18:551–60.

49. Kim S, Parks CG, DeRoo LA, Chen H, Taylor JA, Cawthon RM, et al. Obesity and Weight Gain in Adulthood and Telomere Length. Cancer Epidemiol Biomarkers Prev. 2009;18:816–20.

50. Strandberg TE, Saijonmaa O, Tilvis RS, Pitkälä KH, Strandberg AY, Miettinen TA, et al. Association of telomere length in older men with mortality and midlife body mass index and smoking. J Gerontol A Biol Sci Med Sci. 2011;66:815–20.

51. Moreno-Navarrete JM, Ortega F, Sabater M, Ricart W, Fernandez-Real JM. Telomere length of subcutaneous adipose tissue cells is shorter in obese and formerly obese subjects. Int J Obes. 2010;34:1345–8.

52. Nordfjäll K, Eliasson M, Stegmayr B, Lundin S, Roos G, Nilsson PM. Increased abdominal obesity, adverse psychosocial factors and shorter telomere length in subjects reporting early ageing; the MONICA Northern Sweden Study. Scand J Public Health. 2008;36:744–52.

53. Zannolli R, Mohn A, Buoni S, Pietrobelli A, Messina M, Chiarelli F, et al. Telomere length and obesity. Acta Pædiatrica. 2008;97:952–4.

54. Chen S, Yeh F, Lin J, Matsuguchi T, Blackburn E, Lee ET, et al. Short leukocyte telomere length is associated with obesity in American Indians: the Strong Heart Family study. Aging. 2014;6:380–9.

55. Buxton JL, Das S, Rodriguez A, Kaakinen M, Couto Alves A, Sebert S, et al. Multiple measures of adiposity are associated with mean leukocyte telomere length in the northern Finland birth cohort 1966. PloS One. 2014;9:e99133.

56. Njajou OT, Cawthon RM, Blackburn EH, Harris TB, Li R, Sanders JL, et al. Shorter telomeres are associated with obesity and weight gain in the elderly. Int J Obes 2005. 2012;36:1176–9.

57. Diaz VA, Mainous AG, Player MS, Everett CJ. Telomere length and adiposity in a racially diverse sample. Int J Obes 2005. 2010;34:261–5.

58. Du M, Prescott J, Cornelis MC, Hankinson SE, Giovannucci E, Kraft P, et al. Genetic predisposition to higher body mass index or type 2 diabetes and leukocyte telomere length in the Nurses' Health Study. PloS One. 2013;8:e52240.

59. Schaefer C, Sciortino S, Kvale M, Lapham K, Ranatunga D, Rowell S, et al. B4-3: Demographic and Behavioral Influences on Telomere Length and Relationship with All-Cause Mortality: Early Results from the Kaiser Permanente Research Program on Genes, Environment, and Health (RPGEH). Clin Med Res. 2013;11:146–146.

60. Centers for Disease Control and Prevention (CDC. Smoking-attributable mortality, years of potential life lost, and productivity losses--United States, 2000-2004. MMWR Morb Mortal Wkly Rep. 2008;57:1226.

61. Carty CL, Kooperberg C, Liu J, Herndon M, Assimes T, Hou L, et al. Leukocyte Telomere Length and Risks of Incident Coronary Heart Disease and Mortality in a Racially Diverse Population of Postmenopausal Women. Arterioscler Thromb Vasc Biol. 2015;35:2225–31.

62. Aviv A, Valdes A, Gardner JP, Swaminathan R, Kimura M, Spector TD. Menopause Modifies the Association of Leukocyte Telomere Length with Insulin Resistance and Inflammation. J Clin Endocrinol Metab. 2006;91:635–40.

63. O'Donnell CJ, Demissie S, Kimura M, Levy D, Gardner JP, White C, et al. Leukocyte Telomere Length and Carotid Artery Intimal Medial Thickness: The Framingham Heart Study. Arterioscler Thromb Vasc Biol. 2008;28:1165–71.

64. Sun Q, Shi L, Prescott J, Chiuve SE, Hu FB, De Vivo I, et al. Healthy Lifestyle and Leukocyte Telomere Length in U.S. Women. PLoS ONE. 2012;7:e38374.

65. Huzen J, Wong LSM, van Veldhuisen DJ, Samani NJ, Zwinderman AH, Codd V, et al. Telomere length loss due to smoking and metabolic traits. J Intern Med. 2014;275:155–63.

66. Bendix L, Thinggaard M, Fenger M, Kolvraa S, Avlund K, Linneberg A, et al. Longitudinal Changes in Leukocyte Telomere Length and Mortality in Humans. J Gerontol A Biol Sci Med Sci. 2014;69A:231–9.

67. Weischer M, Bojesen SE, Nordestgaard BG. Telomere shortening unrelated to smoking, body weight, physical activity, and alcohol intake: 4,576 general population individuals with repeat measurements 10 years apart. PLoS Genet. 2014;10.

68. Müezzinler A, Mons U, Dieffenbach AK, Butterbach K, Saum K-U, Schick M, et al. Smoking habits and leukocyte telomere length dynamics among older adults: Results from the ESTHER cohort. Exp Gerontol. 2015;70:18–25.

69. Bischoff C, Graakjaer J, Petersen HC, Hjelmborg J v B, Vaupel JW, Bohr V, et al. The heritability of telomere length among the elderly and oldest-old. Twin Res Hum Genet Off J Int Soc Twin Stud. 2005;8:433–9.

70. Bekaert S, De Meyer T, Rietzschel ER, De Buyzere ML, De Bacquer D, Langlois M, et al. Telomere length and cardiovascular risk factors in a middle-aged population free of overt cardiovascular disease. Aging Cell. 2007;6:639–47.

71. Fitzpatrick AL, Kronmal RA, Gardner JP, Psaty BM, Jenny NS, Tracy RP, et al. Leukocyte Telomere Length and Cardiovascular Disease in the Cardiovascular Health Study. Am J Epidemiol. 2007;165:14–21.

72. Brouilette SW, Moore JS, McMahon AD, Thompson JR, Ford I, Shepherd J, et al. Telomere length, risk of coronary heart disease, and statin treatment in the West of Scotland Primary Prevention Study: a nested case-control study. The Lancet. 13;369:107–14.

73. Corrao G, Bagnardi V, Zambon A, La Vecchia C. A meta-analysis of alcohol consumption and the risk of 15 diseases. Prev Med. 2004;38:613–9.

74. Di Castelnuovo A, Costanzo S, Bagnardi V, Donati MB, Iacoviello L, de Gaetano G. Alcohol dosing and total mortality in men and women: an updated meta-analysis of 34 prospective studies. Arch Intern Med. 2006;166:2437–45.

75. Ferreira MP, Willoughby D. Alcohol consumption: the good, the bad, and the indifferent. Appl Physiol Nutr Metab. 2008;33:12–20.

76. Strandberg TE, Strandberg AY, Saijonmaa O, Tilvis RS, Pitkälä KH, Fyhrquist F. Association between alcohol consumption in healthy midlife and telomere length in older men. The Helsinki Businessmen Study. Eur J Epidemiol. 2012;27:815–22.

77. Romano GH, Harari Y, Yehuda T, Podhorzer A, Rubinstein L, Shamir R, et al. Environmental Stresses Disrupt Telomere Length Homeostasis. PLoS Genet. 2013;9:e1003721.

78. Warburton DER, Nicol CW, Bredin SSD. Health benefits of physical activity: the evidence. CMAJ Can Med Assoc J J Assoc Medicale Can. 2006;174:801–9.

79. RADÁK Z, CHUNG HY, NAITO H, TAKAHASHI R, JUNG KJ, KIM H-J, et al. Age-associated increase in oxidative stress and nuclear factor κB activation are attenuated in rat liver by regular exercise. FASEB J. 2004;18:749–50.

80. Cherkas LF, Hunkin JL, Kato BS, Richards JB, Gardner JP, Surdulescu GL, et al. The association between physical activity in leisure time and leukocyte telomere length. Arch Intern Med. 2008;168:154–8.

81. Du M, Prescott J, Kraft P, Han J, Giovannucci E, Hankinson SE, et al. Physical activity, sedentary behavior, and leukocyte telomere length in women. Am J Epidemiol. 2012;175:414–22.

82. Hosgood III HD, Cawthon R, He X, Chanock S, Lan Q. Genetic variation in telomere maintenance genes, telomere length, and lung cancer susceptibility. Lung Cancer. 2009;66:157–61.

83. Jang JS, Choi YY, Lee WK, Choi JE, Cha SI, Kim YJ, et al. Telomere length and the risk of lung cancer. Cancer Sci. 2008;99:1385–9.

84. 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;95:1211–8.

85. Ma H, Zhou Z, Wei S, Liu Z, Pooley KA, Dunning AM, et al. Shortened Telomere Length Is Associated with Increased Risk of Cancer: A Meta-Analysis. PLoS ONE. 2011;6:e20466.

86. Martinez-Delgado B, Yanowsky K, Inglada-Perez L, de la Hoya M, Caldes T, Vega A, et al. Shorter telomere length is associated with increased ovarian cancer risk in both familial and sporadic cases. J Med Genet. 2012;49:341–4.

87. Terry KL, Tworoger SS, Vitonis AF, Wong J, Titus-Ernstoff L, De Vivo I, et al. Telomere Length and Genetic Variation in Telomere Maintenance Genes in Relation to Ovarian Cancer Risk. Cancer Epidemiol Biomarkers Prev. 2012;21:504–12.

88. Qin, Q, Sun, J, Yin, J, Liu, L, Chen, J, Zhang, Y, et al. Telomere Length in Peripheral Blood Leukocytes Is Associated with Risk of Colorectal Cancer in Chinese Population. PLoS ONE. 2014;9:e88135.

89. Shen J, Gammon MD, Terry MB, Wang Q, Bradshaw P, Teitelbaum SL, et al. Telomere length, oxidative damage, antioxidants and breast cancer risk. Int J Cancer. 2009;124:1637–43.

90. Qu S, Wen W, Shu X-O, Chow W-H, Xiang Y-B, Wu J, et al. Association of Leukocyte Telomere Length With Breast Cancer Risk: Nested Case-Control Findings From the Shanghai Women's Health Study. Am J Epidemiol. 2013;177:617–24.

91. Shen J, Terry MB, Gurvich I, Liao Y, Senie RT, Santella RM. Short Telomere Length and Breast Cancer Risk: A Study in Sister Sets. Cancer Res. 2007;67:5538–44.

92. Lan Q, Cawthon R, Gao Y, Hu W, Hosgood HD III, 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:e59230.

93. Zheng Y-L, Ambrosone C, Byrne C, Davis W, Nesline M, McCann S. Telomere length in blood cells and breast cancer risk: investigations in two case–control studies. Breast Cancer Res Treat. 2010;120:769–75.

94. Pooley KA, Sandhu MS, Tyrer J, Shah M, Driver KE, Luben RN, et al. Telomere Length in Prospective and Retrospective Cancer Case-Control Studies. Cancer Res. 2010;70:3170–6.

95. Svenson U, Nordfjäll K, Stegmayr B, Manjer J, Nilsson P, Tavelin B, et al. Breast Cancer Survival Is Associated with Telomere Length in Peripheral Blood Cells. Cancer Res. 2008;68:3618–23.

96. De Vivo I, Prescott J, Wong JYY, Kraft P, Hankinson SE, Hunter DJ. A Prospective Study of Relative Telomere Length and Postmenopausal Breast Cancer Risk. Cancer Epidemiol Biomarkers Prev. 2009;18:1152–6.

97. Zee RYL, Castonguay AJ, Barton NS, Buring JE. Mean Telomere Length and Risk of Incident Colorectal Carcinoma: A Prospective, Nested Case-Control Approach. Cancer Epidemiol Biomarkers Prev. 2009;18:2280–2.

98. Cui Y, Cai Q, Qu S, Chow W-H, Wen W, Xiang Y-B, et al. Association of Leukocyte Telomere Length with Colorectal Cancer Risk: Nested Case–Control Findings from the Shanghai Women's Health Study. Cancer Epidemiol Biomarkers Prev. 2012;21:1807–13.

99. Wentzensen IM, Mirabello L, Pfeiffer RM, Savage SA. The Association of Telomere Length and Cancer: a Meta-analysis. Cancer Epidemiol Biomarkers Prev. 2011;20:1238–50.

100. Rode L, Bojesen SE, Weischer M, Vestbo J, Nordestgaard BG. Short telomere length, lung function and chronic obstructive pulmonary disease in 46,396 individuals. Thorax. 2013;68:429–35.

101. Allshire RC, Dempster M, Hastie ND. Human telomeres contain at least three types of G–rich repeat distributed non-randomly. Nucleic Acids Res. 1989;17:4611–27.

102. Cawthon RM. Telomere measurement by quantitative PCR. Nucleic Acids Res. 2002;30:e47–e47.

103. Burgess S, Butterworth A, Thompson SG. Mendelian Randomization Analysis With Multiple Genetic Variants Using Summarized Data. Genet Epidemiol. 2013;37:658–65.

CHAPTER 2: GENETIC DETERMINANTS OF TELOMERE LENGTH AND RISK OF COMMON CANCERS: A MENDELIAN RANDOMIZATION STUDY*

Chenan Zhang[1], Jennifer A. Doherty[2], Stephen Burgess[3], Rayjean J. Hung[4], Sara Lindström[5], Peter Kraft[5], Jian Gong[6], Christopher Amos[7], Thomas A. Sellers[8], Alvaro N.A. Monteiro[8], Georgia Chenevix-Trench[9], Heike Bickeböller[10], Angela Risch[11,12], Paul Brennan[13], James McKay[13], Richard Houlston[14], Maria Teresa Landi[15], Maria Timofeeva[13], Yufei Wang[14], Joachim Heinrich[16], Zsofia Kote-Jarai[17], Rosalind A Eeles[17,18], Ken Muir[19.20], Fredrik Wiklund[21], Henrik Grönberg[21], Sonja I. Berndt[15], Stephen J. Chanock[15], Fredrick Schumacher[22], Christopher A. Haiman[22], Brian E. Henderson[22], Ali Amin Al Olama[23], Irene L. Andrulis[24], John L. Hopper[25], Jenny Chang-Claude[26], Esther M. John[27,30], Kathleen E. Malone[6], Marilie D. Gammon[28], Giske Ursin[29], Alice S. Whittemore[30], David J. Hunter[5], Stephen B. Gruber[31],Julia A. Knight[32-34], Lifang Hou[35], Loic Le Marchand[36], Polly A. Newcomb[6,37], Thomas J. Hudson[38], Andrew T. Chan[39,40], Li Li[41], Michael O. Woods[42], Habibul Ahsan[1,43-45], Brandon L. Pierce[1,43,45]*, on behalf of GECCO and the GAME-ON Network: CORECT, DRIVE, ELLIPSE, FOCI, and TRICL

---

*This is a pre-copyedited, author-produced PDF of an article accepted for publication in Human Molecular Genetics following peer review. The version of record (Human Molecular Genetics, 2015, Vol. 24, No. 18 5356–5366) is available online at: hmg.oxfordjournals.org/content/24/18/5356.

Institutions:

1) Department of Public Health Sciences, The University of Chicago, Chicago, IL, USA; 2) Section of Biostatistics & Epidemiology, The Geisel School of Medicine at Dartmouth, Lebanon, NH, USA; 3) Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK; 4) Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital, Toronto, Canada; 5) Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA; 6) Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA; 7) Center for Genomic Medicine, Department of Community and Family Medicine, Geisel School of Medicine, Dartmouth College, Lebanon, NH, USA; 8) Department of Cancer Epidemiology, H. Lee Moffitt Cancer Center, Tampa, FL, USA; 9) Department of Genetics, QIMR Berghofer Medical Research Institute, Brisbane, Australia; 10) Department of Genetic Epidemiology, University Medical Center, Georg-August-University Göttingen, Göttingen, Germany; 11) DKFZ - German Cancer Research Center, Translational Lung Research Centre Heidelberg (TLRC-H), Member of the German Center for Lung Research, Heidelberg, Germany; 12) Translational Lung Research Centre Heidelberg (TLRC-H), Member of the German Center for Lung Research, Heidelberg, Germany; 13) International Agency for Research on Cancer, Lyon, France; 14) Division of Genetics and Epidemiology, Institute of Cancer Research, Sutton, Surrey, UK; 15) Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, U.S. Public Health Service, Bethesda, MD, USA; 16) Institute of Epidemiology I, Helmholtz Zentrum München, Neuherberg, Germany; 17) The Institute of Cancer Research, Sutton, UK; 18) Royal Marsden National Health Service (NHS) Foundation Trust, London and Sutton, UK; 19) Warwick Medical School, University of Warwick, Coventry, UK; 20) Institute of Population Health, University of Manchester, Manchester, UK; 21)

Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm, Sweden; 22) Department of Preventive Medicine, Keck School of Medicine, University of Southern California/Norris Comprehensive Cancer Center, Los Angeles, California, USA; 23) Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK; 24) Molecular Genetics/Laboratory Medicine and Pathobiology, Mount Sinai Hospital, University of Toronto, Toronto, Canada; 25) Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Parkville, Australia; 26) Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany; 27) Cancer Prevention Institute of California, Fremont, CA, USA; 28) Department of Epidemiology, University of North Carolina School of Public Health, Chapel Hill, NC, USA; 29) Kreftregisteret, Cancer Registry of Norway, Oslo, Norway; 30) Stanford University School of Medicine, Stanford, CA, USA; 31) USC Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, CA, USA; 32) Ontario Cancer Genetics Network, Fred A. Litwin Center for Cancer Genetics, Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Ontario, Canada; 33) Division of Epidemiology, Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada; 34) Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Ontario, Canada; 35) Department of Preventive Medicine, Northwestern University, Chicago, IL, USA; 36) Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI, USA; 37) Department of Epidemiology, University of Washington School of Public Health, Seattle, WA, USA; 38) Ontario Institute for Cancer Research, Toronto, Ontario, Canada; University of Toronto, Toronto, Ontario, Canada; 39) Division of Gastroenterology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA; 40) Channing Division of Network

Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA; 41) Department of Family Medicine and Community Health, Case Western Reserve University, Cleveland, OH, USA; 42) Discipline of Genetics, Faculty of Medicine, Memorial University, Newfoundland & Labrador, Canada; 43) Center for Cancer Epidemiology and Prevention, The University of Chicago, Chicago, IL, USA; 44) Department of Medicine, The University of Chicago, Chicago, IL, USA; 45) Department of Human Genetics, The University of Chicago, Chicago, IL, USA

ABSTRACT

Epidemiological studies have reported inconsistent associations between telomere length (TL) and risk for various cancers. These inconsistencies are likely attributable, in part, to biases that arise due to post-diagnostic and post-treatment TL measurement. To avoid such biases, we used a Mendelian randomization approach and estimated associations between nine TL-associated SNPs and risk for five common cancer types (breast, lung, colorectal, ovarian and prostate cancer, including subtypes) using data on 51,725 cases and 62,035 controls. We then used an inverse-variance weighted average of the SNP-specific associations to estimate the association between a genetic score representing long TL and cancer risk. The long TL genetic score was significantly associated with increased risk of lung adenocarcinoma (P=6.3x10^{-15}), even after exclusion of a SNP residing in a known lung cancer susceptibility region (*TERT-CLPTM1L*) P=6.6x10^{-6}). Under Mendelian randomization assumptions, the association estimate (odds ratio (OR)=2.78) is interpreted as the OR for lung adenocarcinoma corresponding to a 1000 base pair increase in TL. The weighted TL SNP score was not associated with other cancer types or subtypes. Our finding that genetic determinants of long TL increase lung adenocarcinoma risk avoids issues with reverse causality and residual confounding that arise in observational studies of TL and disease risk. Under Mendelian randomization assumptions, our finding suggests that longer TL increases lung adenocarcinoma risk. However, caution regarding this causal interpretation is warranted in light of the potential issue of pleiotropy, and a more general interpretation is that SNPs influencing telomere biology are also implicated in lung adenocarcinoma risk.

INTRODUCTION

Telomeres are DNA-protein complexes at chromosome ends that help maintain genome stability by protecting DNA from damage and fusion. The DNA component is a 6-base TTAGGG repeat sequence that shortens with each cell division. In differentiated cells, telomere shortening eventually leads to loss of telomere protection and genome instability, typically triggering cell senescence or programmed cell death (1). In stem and progenitor cells, the telomerase enzyme elongates telomeres, enabling prolonged cell survival (2). Telomerase is also reactivated in >90% of human tumors (3), which typically have short telomeres (a potential cause of genome instability), thus promoting proliferation and survival (4).

The critical role of telomeres and telomerase in carcinogenesis has led to the hypothesis that short telomere length (TL) is a risk factor for cancer (5). Indeed, short relative TL measured in surrogate tissues, such as peripheral blood cells, has been associated with increased risk for lung (6, 7), ovarian (8), colorectal (9), and breast cancers (10, 11) in epidemiological studies (with the interpretation that blood TL predicts cancer risk because it is a proxy for TL in cancer-prone tissues). However, such associations are not consistent across all cancers or even within cancer types, with some studies reporting null, U-shaped, or positive associations (11–16). Furthermore, due to the retrospective nature of case-control studies from which many of these association estimates are obtained, telomere shortening that occurs after diagnosis, potentially due to treatment (17, 18) or disease progression, can result in biased estimates of the association between TL and cancer risk (6, 8, 14).

Genome-wide association (GWA) studies have identified several genomic regions containing variants associated with TL (19–21), including the *TERT* (telomerase reverse transcriptase) region on 5p15.33. Furthermore, GWA studies of cancer risk have observed that

variants in the *TERT* region influence risk for multiple cancer types, including breast (22), colorectal (23), lung (24), prostate (24), and ovarian (22) cancer, although these associations do not appear to all be driven by the same causal variant. In light of this evidence indicating an important role for telomeres in carcinogenesis, we undertook a comprehensive examination of associations between genetic determinants of TL and cancer risk.

In this work, we describe the associations between nine TL-associated genetic variants and risk for five cancer types (breast, lung, colorectal, ovarian and prostate), using data from the Genetic Associations and Mechanisms in Oncology (GAME-ON) network of consortia for post-GWA research. In addition, we estimate the association between a multi-variant TL score and cancer risk, which corresponds to the effect of TL on cancer risk under Mendelian randomization assumptions (25). However, this interpretation requires caution because the validity of the Mendelian randomization assumptions (such as the absence of pleoitropy) cannot be proven. Because genotype-phenotype associations are not vulnerable to biases caused by reverse causation or confounding by environment, the Mendelian randomization approach used in this study is an attractive approach for estimating relationships between TL and cancer risk that is not prone to biases due to confounding by environment, lifestyle, or treatment-related factors.

RESULTS

The Genetic Associations and Mechanisms in Oncology (GAME-ON) Consortium is a network of five consortia focused on cancers of the breast, colon, lung, ovary, and prostate. The GAME-ON network represents 33 GWA studies and is comprised of >51,000 cancer cases and >62,000 controls (26). Samples sizes for each cancer type and subtype are listed in Table 2.1.

Table 2.1. Sample sizes for cancer types included in the Genetic Associations and Mechanisms in Oncology (GAME-ON) consortium. Details on the GAME-ON Network and the contributing GWA studies have been previously described (26): (http://epi.grants.cancer.gov/gameon/)

Cancer Type	Cases	Controls	GWA studies[a]
Breast			
All	15,748	18,084	11
ER-negative	4,939	13,128	8
Colorectal	5,100	4,831	6
Lung[b]			
All	12,160	16,838	9
Adenocarcinoma	3,718	15,871	9
Squamous	3,422	16,015	9
Ovarian[b]			
All	4,369	9,123	3
Clear-cell	356	9,123	3
Endometrioid	715	9,123	3
Serous	2,556	9,123	3
Prostate			
All	14,160	12,724	6
Aggressive	4,450	12,724	6

[a]Not including studies from the Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO)
[b]Subtypes not representative of all existing subtypes within cancer type

Association estimates for individual SNPs

Based on the existing literature, we identified nine SNPs showing genome-wide significant associations ($P<5x10^{-8}$) with TL in GWA studies (19–21). From these prior papers we obtained the identifier for the lead SNP at each reported locus as well as the "long TL" allele, association estimate for the "long" allele (in terms of kb of TL per allele), standard error, and P-value for each SNP's association with TL (Table 2.2). We estimated associations between each of the nine TL-associated SNPs and risk for each of the five common cancer types and subtypes in the GAME-ON study, shown as forest plots in Supplementary Figure 2.1. Of note, for all nine

Table 2.2. Characteristics of genetic variants associated with relative telomere length as reported in prior genome-wide association studies

SNP Identifier	Chr.	Locus	"Long" Allele	Beta estimate[a]	P-value	Source
rs10936599	3	TERC	C	0.117	$2.5x10^{-31}$	Codd et al., 2013(19)
rs2736100	5	TERT	C	0.094	$4.4x10^{-19}$	Codd et al., 2013(19)
rs7726159[b]	5	TERT	A	0.073	$4.7x10^{-17}$	Pooley et al., 2013(21)
rs7675998	4	NAF1	G	0.090	$4.3x10^{-16}$	Codd et al., 2013(19)
rs9420907	10	OBFC1	C	0.083	$6.9x10^{-11}$	Codd et al., 2013(19)
rs6772228	3	PXK	T	0.120	$3.9x10^{-10}$	Pooley et al., 2013(21)
rs8105767	19	ZNF208	G	0.058	$1.1x10^{-9}$	Codd et al., 2013(19)
rs755017	20	RTEL1	G	0.074	$6.7x10^{-9}$	Codd et al., 2013(19)
rs412658[c]	19	ZNF676	T	0.050	$9.8x10^{-9}$	Mangino et al., 2012(20)
rs3027234	17	CTC1	C	0.057	$2.3x10^{-8}$	Mangino et al., 2012(20)
rs11125529	2	ACYP2	A	0.067	$4.5x10^{-8}$	Codd et al., 2013(19)

[a]Reported in kb telomere per "long" allele
[b]In linkage disequilibrium (r^2=0.382) with rs2736100 of the *TERT* locus, excluded from all analyses
[c]In linkage disequilibrium(r^2=0.704) with rs8105767 of the *ZNF208* locus, excluded from all analyses

SNPs, the long TL allele had an OR >1 for lung adenocarcinoma, with four of the nine associations being statistically significant (P<0.05) (Figure 2.1, top left). In contrast, no TL-associated SNP was significantly associated with squamous cell carcinoma of the lung (Figure 2.1, bottom left). Prostate cancer risk also showed nominally significant positive associations with the long TL alleles for three of the nine SNPs (P<0.05) (Supplementary Figure 2.1).

Mendelian randomization estimates based on multi-SNP scores

We estimated the associations between a "multi-SNP TL score" and risk for each cancer (Table 2.3) using a previously-described Mendelian randomization approach that obtains an estimate using an inverse-variance weighted average of SNP-specific associations (27, 28) (see Methods). Of note, we identified a highly statistically significant association between long TL and increased risk of lung adenocarcinoma with an odds ratio (OR) of 2.78 per 1 kb increase TL (95% Confidence Interval (CI) 2.16, 3.58; $P = 6.3x10^{-15}$). However, we observed no such

association for squamous cell carcinoma of the lung. Associations for these two lung cancer subtypes are displayed in Figure 2.1 (right) as solid red lines (slope=ln(OR)) overlaid on the association estimates for the nine SNPs that were used to generate the OR. A positive slope indicates that longer TL is associated with increased cancer risk, while a negative slope indicates that longer TL is associated with decreased risk. The correlation (r) between the magnitude of the SNPs' associations with TL and the magnitude of the SNPs' associations with adenocarcinoma risk was 0.17. Other than lung adenocarcinoma, no other cancer types showed a statistically significant association with the multi-SNP score. However, prostate cancer risk showed suggestive evidence of positive association with long TL with a Mendelian randomization OR of 1.21 per 1 kb increase in TL (95% CI 0.99, 1.46; $P = 0.06$). Scatter plots for all cancer types are displayed in Supplementary Figure 2.2.

Additional age- and sex-stratified analyses were conducted for overall lung cancer, with findings indicating similar estimates for younger subjects (≤50 years old) (OR = 1.95; 95% CI 1.19, 3.21; $P = 0.008$) and older subjects (>50 years old) (OR = 1.78; 95% CI 1.47, 2.15; $P = 2.98\text{x}10^{-9}$) and similar estimates for men (OR = 1.72; 95% CI 1.38, 2.14; $P = 1.06\text{x}10^{-6}$) and women (OR = 1.99; 95% CI 1.42, 2.77; $P = 5.3\text{x}10^{-5}$).

In addition to the inverse-variance weighted approach for obtaining Mendelian randomization estimates, we also used a likelihood-based Mendelian randomization method (28). Both methods produced very similar estimates for all cancer types, although the lung adenocarcinoma estimates varied more between the two methods compared to the other cancer types (Table 2.3).

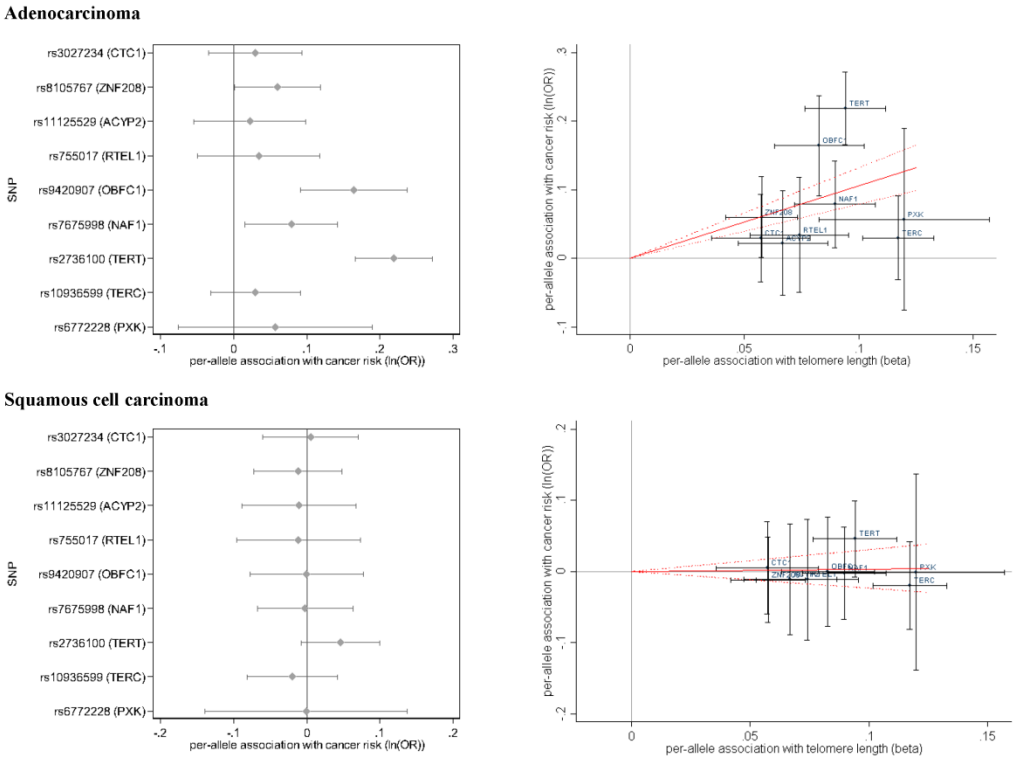
Table 2.3. Odds ratios (ORs) of cancer risk per 1000 base pair increase in telomere length according to a multi-SNP telomere length score using the inverse-variance weighted method (left), and the likelihood method (right)

	Inverse-Variance Weighted Method			Likelihood Method		
Cancer Type	OR	95% CI	P-value	OR	95% CI	P-value
Breast						
All	1.02	0.86, 1.21	0.82	1.02	0.86, 1.21	0.81
ER-negative	1.05	0.81, 1.38	0.70	1.05	0.80, 1.38	0.70
Colorectal	1.25	0.92, 1.69	0.15	1.26	0.92, 1.71	0.15
Lung						
All	1.65	1.39, 1.96	1.3×10^{-8}	1.67	1.40, 2.00	1.3×10^{-8}
Adenocarcinoma	2.87	2.20, 3.74	6.3×10^{-15}	3.03	2.29, 4.01	8.2×10^{-15}
Squamous	1.04	0.79, 1.36	0.79	1.04	0.79, 1.36	0.79
Ovarian						
All	1.13	0.87, 1.47	0.37	1.13	0.87, 1.48	0.36
Clear-cell	1.65	0.78, 3.51	0.19	1.68	0.78, 3.61	0.19
Endometrioid	1.30	0.75, 2.24	0.35	1.30	0.75, 2.25	0.35
Serous	1.19	0.86, 1.65	0.30	1.19	0.86, 1.66	0.29
Prostate						
All	1.21	0.99, 1.46	0.06	1.22	1.00, 1.48	0.06
Aggressive	1.10	0.83, 1.45	0.52	1.10	0.83, 1.46	0.51

Sensitivity Analyses

The estimates reported above can only be interpreted as the causal effect of average TL on cancer risk when the Mendelian randomization assumptions are valid—namely, when 1) the SNPs from the literature are truly predictive of TL in the cancer-prone tissue, 2) the SNPs are not associated with other factors (confounders) that influence both TL and cancer risk, and 3), the SNPs only affect cancer risk through their effects on TL, i.e. there are no alternative causal pathways by which the SNPs influence cancer risk. Violation of any of the assumptions can

Figure 2.1. Forest plots (left) and scatter plots (right) of associations between telomere length-associated SNPs and risk for lung adenocarcinoma (top) and squamous cell carcinoma (bottom). Forest plots show association estimates (with horizontal bars indicating 95% confidence interval) for the "long telomere" allele of each SNP with cancer risk. SNPs are ordered by increasing magnitude of association with telomere length. Scatter plots show the per-allele association with cancer risk plotted against the per-allele association with kb of telomere length (with vertical and horizontal black lines showing 95% confidence interval for each SNP). The scatter plot is overlaid with the Mendelian randomization estimate (slope of red solid line with dotted lines showing 95% confidence interval) of the effect of telomere length on cancer risk



result in a biased causal estimate for the effect of TL on cancer risk. We performed sensitivity analyses in which SNPs were excluded from the multi-SNP score based on potential violation of these assumptions.

To assess a potential violation of the first assumption, an additional analysis was performed after excluding the SNP near the *PXK* region (rs6772228), which may be a false positive association evidenced by its lack of plausible biological explanation, and the lack of consistency in its association with TL across several study sites (21). This "strict" analysis resulted in a notable difference only for prostate cancer risk, which now showed a statistically significant estimate (OR 1.27; 95% CI 1.03, 1.54; $P = 0.02$) (Supplementary Table 2.1).

To assess potential violation of the second and third assumptions, an additional analysis was performed including only SNPs for which the magnitude of association with cancer risk appeared to be proportional to the magnitude of association with TL. Under Mendelian Randomization assumptions two and three, the ratio of the SNP-cancer association to the SNP-TL association for each SNP should be similar for all SNPs used in the SNP score. An inflation of the ratio for any single SNP may be an indication that the SNP exerts a pleiotropic effect on cancer that is unrelated to its effect on TL. Deviation from this expectation is tested using a goodness-of-fit test (29, 30), in which SNPs that exhibit evidence of pleiotropy due to an inflated SNP-cancer to SNP-TL ratio can be detected and excluded from Mendelian randomization analyses (see Materials and Methods).

The exclusion of SNPs that failed the goodness-of-fit test (Supplementary Table 2.2) resulted in two notable findings (Supplementary Table 2.1). First, exclusion of the SNPs in the *TERT* and *CTC1* regions from the SNP set used for overall prostate cancer resulted in a stronger, statistically significant association (OR = 1.45; 95% CI 1.18, 1.82; $P = 7.9 \times 10^{-4}$). Second, the

exclusion of the SNP in the *TERT* region from the SNP set used for lung adenocarcinoma resulted in a somewhat attenuated association, but the association remained statistically significant: (OR = 2.00, 95% CI: 1.48, 2.70, $P = 6.6 \times 10^{-6}$). Lung adenocarcinoma had a significant heterogeneity test statistic (Supplementary Table 2.2), which likely explains why there was a difference in the estimates obtained using the inverse-variance weighted method and the likelihood method noted earlier; the optimization algorithm of the likelihood-based method can have poor convergence when the heterogeneity statistic is strongly significant (28). This difference in estimates between methods is eliminated after excluding the *TERT* SNP (rs2736100) that drives the heterogeneity in association estimates for lung adenocarcinoma.

DISCUSSION

In this analysis of cancer risk across five cancer-prone organs, we observed that a multi-SNP score for long telomeres is significantly associated with increased risk of lung adenocarcinoma (but not squamous cell carcinoma) and suggestively associated with increased risk of prostate cancer. We did not observe an association between the multi-SNP score and risk of breast, colorectal, or ovarian cancer (including subtypes). Under Mendelian randomization assumptions, the associations reported here can be interpreted as effects of TL on cancer risk, although caution regarding such an interpretation is warranted because the validity of these assumptions (such as the absence of pleiotropy) cannot be proven. Our results were consistent using two different analytic approaches. After performing sensitivity analyses in which SNPs were excluded from the multi-SNP score based on potential violation of Mendelian randomization assumptions, the association with prostate cancer risk became statistically significant. In addition, the exclusion of the SNP in the *TERT* region (a known susceptibility

locus for lung cancer) from the lung adenocarcinoma analysis resulted in an attenuated but still highly statistically significant association, indicating that the observed association is not solely driven by the SNP in the *TERT* region. Even after dropping the three SNPs showing significant association with lung adenocarcinoma (in *TERT*, *OBFC1*, and *NAF1*) the association is still nominally significant (OR = 1.54, $P = 0.018$).

Several epidemiologic studies have examined the association between leukocyte TL and lung cancer risk. Three retrospective case-control studies reported an association between long TL and decreased lung cancer risk in US and Korean subjects (7, 31, 32). However, a fourth stratified retrospective case-control study showed a positive association between TL and adenocarcinoma risk, but an inverse association for squamous cell carcinoma (33). In two studies with prospective TL measurement, long TL was found to be associated with increased overall lung cancer risk among Caucasian male smokers (34) and East Asian female never-smokers (13), while a large Danish general population study found no association (12). In a pooled analysis of three prospective cohort studies including the aforementioned two prospective studies and a third study conducted in the United States, long telomeres were associated with increased lung cancer risk, and the association was present in adenocarcinoma while absent in squamous cell carcinoma (35). Consistent with findings from the prospective studies and the stratified case-control study, we observed a positive association between the long TL SNP score and lung cancer risk corresponding to an OR of 1.65 per 1000 base pairs (bp) TL for overall lung cancer ($P = 1.3 \times 10^{-8}$) and an OR of 2.87 per 1000 bp TL for lung adenocarcinoma ($P = 6.3 \times 10^{-15}$) (assuming valid Mendelian randomization assumptions). Seow et al. reported the risk of overall lung cancer between the lowest quartile and highest quartile as OR = 1.86 (95% CI 1.33-2.62), and the risk of adenocarcinoma cancer as OR = 2.52 (95% CI 1.38-4.60) (35). While our

estimates are not directly comparable to these prior estimates due to differences in the scale of the TL variable (kb vs. quartiles), we used a simple simulation to show that our estimates are similar to these prior estimates. We simulated normally distributed TL variables with a mean of 6000 bp and standard deviations ranging from 400 to 700 bp based on values observed from prior literature (36–38). The difference between the mean values for quartiles one and four ranged between 1018-1781 bp. We then rescaled our ORs to correspond to a difference in 1018-1781 bp rather than 1000 bp. The resulting estimates are OR = 1.68-2.94 for overall lung cancer risk between the lowest and highest quartiles, and OR = 2.92-5.11 for lung adenocarcinoma risk between the lowest and highest quartiles. These are comparable with the estimates reported by Seow et al. The observed heterogeneity in our association estimates is likely due to the two subtypes being biologically distinct, having previously been characterized as having different genetic susceptibilities (39), unique gene expression profiles (40), distinct molecular features (41), and different patterns of chromosomal imbalance (42).

The Asian female non-smokers among whom a TL-lung cancer association was observed (13) were also recently studied to evaluate seven TL-associated SNPs in relation to lung cancer risk (43). Consistent with our findings, the risk score for long TL was associated with an increase in lung cancer. Furthermore, their stratified analyses suggested a stronger association among younger individuals (<60 years old) and significant associations for both adenocarcinoma and squamous cell subtypes. In contrast, our stratified analyses produced similar estimates by age and different estimates by subtype. Potential factors driving these differences in findings may be differences in ancestry or differences in the etiology of subtypes between the two study populations.

A protective effect of short TL on lung cancer risk has a biologically plausible explanation, as short telomeres could protect against cancer by triggering cell senescence or programmed cell death in the presence of functional cell cycle checkpoints and intact apoptotic pathways (44). Conversely, long telomeres may enable additional rounds of cell division, allowing more opportunities for the accumulation of somatic mutations that promote carcinogenesis, resulting in greater susceptibility to malignant transformation (45, 46). The association between long TL SNPs and increased cancer risk has also been previously observed in melanoma cancer (47), with a proposed mechanism being that long telomeres increase the proliferative duration of cells, thus delaying senescence and allowing further mutations to occur (48).

Cigarette smoking is a potential confounder in many epidemiologic studies of TL and cancer due to its correlation with short TL (49). However, our study utilizes genetic variants associated with TL as proxy for actual measured TL, and is therefore not subject to the potential confounding effects of smoking and other exposures. TL-outcome confounders such as smoking would only introduce bias if the genetic variants used in score are also associated with smoking behavior, as it would be a violation of the second Mendelian randomization assumption. Smoking-related violation of Mendelian randomization assumptions (e.g., SNPs that influence TL through smoking behavior) is unlikely due to the lack of an association between the TL SNP score and squamous cell carcinoma of the lung, which would be expected to have a stronger association with smoking-related SNPs than adenocarcinoma (50), and due to the lack of evidence that these SNPs influence smoking behavior based on prior GWA studies.

There are few published studies on TL and prostate cancer risk. In a small retrospective case-control study (51) and two prospective studies (12, 52), no statistically significant

association was observed. In contrast, we observe a suggestive association between long TL and increased risk for prostate cancer, an association that increases in significance in the context of the "goodness-of-fit"-based sensitivity analyses. This finding warrants further investigation.

Results from prior studies of TL and colorectal cancer risk are also inconsistent. An inverse association between TL and colorectal cancer risk was described in two case control studies (9, 14), while a U-shaped association (16) and three null-associations (12, 14, 15) were observed in prospective studies. Consistent with the previous null-findings, our results show no significant association between TL and colorectal cancer risk, despite the inclusion of the *TERC* rs10936599 SNP, which was previously reported to be associated with both increased TL and increased risk of colorectal cancer (53). The null findings were also consistent in an analysis of data from the Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO) (54) (Supplementary Figure 2.3).

Although one prospective study showed evidence for association between long TL and increased breast cancer risk (12), two different meta-analyses of TL and breast cancer risk based on multiple retrospective and prospective studies concluded there was no overall evidence of association (6, 8). Our findings for overall breast cancer are in agreement with these prior null studies. In addition, we observe no association between the TL SNP score and ER-negative breast cancer risk. For ovarian cancer, two prior case-control studies observed an association between longer telomeres and decreased risk (55, 56), one case-control study reported no association (57), while a prospective study also reported no association (12). Our results for overall ovarian cancer, as well as three subtypes, provide no evidence of association with TL-associated SNPs. This lack of association is observed despite the inclusion of SNP rs2736100 located in the *TERT* region, which showed a nominally-significant association with the serous

subtype of ovarian cancer ($P = 0.023$) and is in high LD ($r^2 = 0.8$) with a SNP previously observed to be associated with ovarian cancer (22). Estrogen has been demonstrated in experimental studies to have positive effects on telomerase activity (58), and in epidemiologic studies estrogen has been shown to have a positive association with TL (59). With estrogen as a potential confounder of the association between TL and ER-positive breast and ovarian cancers, it will likely be difficult to parse out the specific role of TL in estrogen-related cancer risk in epidemiological studies. However, Mendelian randomization estimates such as those reported here will not be biased due to confounding by estrogen level. These multi-SNP null findings are similar to what Pooley et. al. also observed while investigating individual TL SNP associations with breast, ovarian, and prostate cancer risks (21).

Although it is not possible to prove the validity of the Mendelian randomization assumptions, it is possible to conduct sensitivity analyses to protect against some potential violations of these assumptions. To address a potential violation of the first assumption—that the SNPs are associated with TL in our study population—we conducted analyses excluding *PXK* SNP rs6772228, whose association with TL has been questioned due to the lack of consistency in its association across several study sites (21). After exclusion, the results were essentially unchanged. For overall prostate cancer, however, the association became statistically significant, lending support to the hypothesis that long TL is associated with an increased risk of prostate cancer.

To address potential violations of the assumption that the SNPs do not have effects on cancer risk independent of their effects on TL, we re-estimated the associations between the TL SNP score and cancer after stepwise removal of potentially-problematic SNPs from the SNP set using a goodness-of-fit test of the proportionality of the SNPs' associations with TL and cancer

risk. These exclusions resulted in some attenuation of the association with lung adenocarcinoma, but did not substantially alter our conclusions. For prostate cancer, the association with the multi-SNP score became statistically significant after excluding *TERT* SNP rs2736100 and *CTC1 SNP* rs3027234. The heterogeneity in association of SNPs in *TERT* and *CTC1* identified by the goodness-of-fit test suggests potential pleiotropic effects of these genetic variants through mechanisms other than TL. The association between *TERT* SNPs and breast and ovarian cancer risks via pathways other than TL has been previously observed (22), and potential extra-telomeric roles have previously been suggested for both telomerase (60) and *CTC1* (61), providing a plausible biological basis for excluding the *TERT* and *CTC1* SNPs from our analysis. It is important to note however, that these secondary sensitivity analyses are data-driven, and are presented as a supplement to the primary analyses that include all nine SNPs.

There are several limitations of this work. The summary-level data did not allow for analyses stratified by covariates of interest such as sex and age. Additionally, our analysis assumed a log-linear association between TL and cancer risk, and the existence of a non-linear (e.g. U-shaped) association may limit our ability to detect an association. Our estimates generated using Mendelian randomization are unbiased only if the SNPs analyzed do not affect cancer risk through causal pathways other than those involving TL. This assumption cannot be proven; however, our confidence in the validity of this assumption is strengthened by the fact that our primary finding is robust to the exclusion of SNPs with potential pleiotropic actions based on prior evidence (*TERT*) and a goodness-of-fit test (Supplementary Tables 2.1, 2) (although it is possible that the goodness-of-fit test is underpowered to identify pleiotropic effects). Our power to detect associations is limited by the small variance in measured TL explained by SNPs used in this analysis (1-2%) (62), although the GAME-ON Network provides

very large sample sizes that enable the detection of strong to moderate associations (Supplementary Table 2.3). Finally, we cannot confirm that genetic determinants of leukocyte TL also predict tissue-specific TL due to the lack of tissue-based TL measures in genome-wide association studies. A potential consequence of selecting SNPs lacking tissue-specific association with TL would be a bias toward the null, reducing our power to detect associations using TL SNP scores. However, correlations between TL measured in blood and TL measured in lung (63, 64) and other tissues (65) have been reported (r=0.35-0.84), consistent with the assumption that SNPs predict TL across multiple tissues. Systematic studies on other tissue types are needed to further address this uncertainty.

In conclusion, in this comprehensive Mendelian randomization study of TL and risk for common cancers, we observed a highly significant association between genetic determinants of long TL and increased risk for lung adenocarcinoma. The estimates reported here are not vulnerable to biases caused by reverse causation or confounding by unmeasured environmental factors, strengthening the evidence for a causal role for telomere length in lung adenocarcinoma. However, the validity of Mendelian randomization estimates is dependent upon several assumptions, namely no pleiotropic effects (independent of TL) of SNPs on the cancer risk or confounders of the TL-cancer relationship. The multi-SNP score for TL should be further investigated as a predictor of adenocarcinoma of the lung, a common lung cancer subtype in both smokers and non-smokers. Future research efforts need to be undertaken to determine the value of telomeres as a novel risk measure or a modifiable pharmacological target, with the long term goal of improving cancer prediction and prevention.

MATERIALS AND METHODS

The GAME-ON network of consortia for post-GWA research

The goals for each individual center of the GAME-ON consortium were to pool data from GWA studies to identify new loci, conduct functional studies in the laboratory to identify causal SNPs and biological mechanisms, and to investigate gene-gene and gene-environment interactions as a part of efforts to develop risk prediction models. A secondary goal was to test hypotheses across the centers that might illuminate common mechanisms of susceptibility. Details of GAME-ON and the participating studies are available at http://epi.grants.cancer.gov/gameon/, and described previously (26).

Identification of SNPs associated with TL

We identified nine SNPs showing independent genome-wide significant associations ($P<5x10^{-8}$) with TL in previously published GWA studies among individuals of European ancestry (19–21). Although there are specific cancer susceptibility regions of interest such as the previously described *TERT* locus, our selection of SNPs is based entirely on the SNPs' ability to predict TL based on prior literature, because predictive accuracy is directly related to statistical power for Mendelian randomization (62). The proportion of variance in measured average TL that is explained by individual SNPs ranges from 0.06% to 0.2% (19), and is up to 1.6% for a combined subset of 4 SNPs (20) (no estimate is currently available for all 9 SNPs). Based on the existing literature, we obtained the identifier for the lead SNP at each reported locus as well as the "long TL" allele, association estimate for the "long" allele (in terms of kb increase in TL per allele), standard error, and P-value for each SNP's association with TL (Table 2.2). Only the lead SNP from each region was included in the analysis. Although the estimates for these 9 SNPs

were obtained from three different studies using two different methods (quantitative PCR and Southern blot of the terminal restriction fragment), we scaled the estimates to the same units (kb of telomere length per risk allele). Comparability between studies is supported by previous studies showing that T/S ratio from qPCR using the Cawthon method is strongly correlated with mean terminal restriction fragment obtained from Southern blot for non-extreme TL values (66). Data on these nine SNPs were available as summary statistics for all cancer types analyzed in the GAME-ON consortium, with the exception of colorectal cancer, for which we obtained proxy SNPs based on strong linkage disequilibrium using the Broad Institute's SNP Annotation and Proxy Search tool (67) (Supplementary Table 2.4).

Statistical analysis

For each cancer type, standard fixed-effects meta-analysis methods were used to combine results from individual GWA studies. For each cancer type, genotyping was performed using Illumina or Affymetrix arrays of varying densities described elsewhere (26). Quality control steps taken include gender identity and chromosomal anomaly check, exclusion of related individuals, principal component-based exclusion of individuals of non-European ancestry, exclusion of SNPs and individuals with substantial missingness, exclusion of SNPs in Hardy-Weinberg disequilibrium, and other sample and SNP quality measures. For each study, imputation was performed using the 1000 Genomes Project Phase 1 version 3 reference haplotypes, resulting in up to ~10 million SNPs being available for the analysis for each cancer type.

Associations between SNPs and cancer risk were estimated using unconditional logistic regression adjusted for age, sex (when applicable), and top principal components (ranging from 2

to 6 across 48 contributing GWA studies). For the lung cancer study, the association was also adjusted for smoking pack-years. We performed the analyses separately for cancer subtypes, including breast (estrogen receptor negative), lung (adenocarcinoma and squamous cell), ovarian (clear cell, endometrioid, and serous), and prostate (aggressive and non-aggressive as previously defined (68)). We performed age- and sex-stratified analyses for overall lung cancer, for which only SNPs imputed to the Illumina 500K array using the HapMap2 reference panel were available across all sites. For the TL SNPs not available on the 500K array, we were able to identify tag SNPs (r2>0.8) for all SNPs except PXK SNP rs6772228 (Supplementary Table 2.5).

We estimated the association between a "multi-SNP TL score" and risk for each cancer using two different Mendelian randomization methods that require only summarized association estimates for each SNP (hence, no actual score is created for each individual, but we estimated the association that would be observed for such as score if individual-level data were used). This approach is appropriate given that the consortium provides only summary estimates rather than individual-level data for SNP associations. The first method is an inverse-variance weighted average of SNP-specific associations that has been described previously (27, 28). The Mendelian randomization estimate $\hat{\beta}_{IVW}$, and its standard error $se(\hat{\beta}_{IVW})$ were calculated using the following equations:

$$\hat{\beta}_{IVW} = \frac{\sum_k X_k Y_k \sigma_{Y_k}^{-2}}{\sum_k X_k^2 \sigma_{Y_k}^{-2}} \qquad (1)$$

$$se(\hat{\beta}_{IVW}) = \sqrt{\frac{1}{\sum_k X_k^2 \sigma_{Y_k}^{-2}}} \qquad (2)$$

where X_k is the per-allele estimate of the kth SNP on telomere length, Y_k is the per-allelle estimate of the SNP on the log-odds of cancer, and σ_{Yk} is the corresponding standard error. A schematic summarizing the aforementioned steps is shown in Supplementary Figure 2.4.

The second method is a likelihood-based method that has been described previously (28). In brief, TL and cancer risk were jointly modeled using a bivariate normal distribution for each genetic variant. The model parameters, including a joint linear effect between TL and cancer log-odds, were estimated using maximum likelihood on the observed data. The likelihood-based analyses were performed using web-based software (http://spark.rstudio.com/sb452/summarized/) (28).

The estimates obtained from the methods described above can be interpreted as the effect of average TL on cancer risk under the following assumptions as previously described for causal inference based on Mendelian randomization (25): 1) the SNPs from the literature are truly predictive of TL in the cancer-prone tissue, 2) the SNPs are not associated with other factors (confounders) that influence both TL and cancer risk, and 3), the SNPs only affect cancer risk through their effects on TL, i.e. there are no alternative causal pathways by which the SNPs influence cancer risk.

To visualize the association results for the SNP score, we plotted the association between each SNP and cancer risk against associations with TL (based on prior literature). Under the assumption that a SNP's association with TL is proportional to its association with cancer risk, one would expect the plotted points to fall along a line that passes through the origin and has a slope equal to the Mendelian randomization estimate. Thus, a steeper slope indicates a stronger magnitude of association between TL and cancer risk. Conversely assuming no causal effect of TL on cancer risk, the Mendelian randomization estimate would correspond to a slope of zero.

To assess a potential violation of the first assumption (i.e., a true association between each of the variant and TL), an additional analysis was performed after excluding the SNP near the *PXK* region (rs6772228), which may be a false positive association evidenced by its lack of plausible biological explanation, and the lack of consistency in its association with TL across several study sites (21). This analysis is referred to as the "strict" analysis. To assess potential violation of the second and third assumptions (i.e., no confounding or pleiotropy), a goodness-of-fit test was performed for each SNP set under the null hypothesis that each SNP included in the SNP score has an association with cancer risk that is proportional to its association with TL. The rejection of the null hypothesis indicated heterogeneity of the associations between SNPs and cancer risk relative to the associations between the SNPs and TL. In instances where the null hypothesis was rejected, stepwise removal of SNPs from the SNP set was performed until there was no significant heterogeneity, based on a method previously described (http://cran.r-project.org/web/packages/gtx/) (29). Specifically, For *K* uncorrelated SNPs,

$$\chi_k^2 = \sum_k Y_k^2 \sigma_{Y_k}^{-2} \qquad (3)$$

$$\chi_{rs}^2 = \left(\frac{\widehat{\beta}_{IVW}}{se(\widehat{\beta}_{IVW})}\right)^2 \qquad (4)$$

in which χ_k^2 and χ_{rs}^2 are $\chi_{(k)}^2$ and $\chi_{(1)}^2$ distributed test statistics, respectively, for the association between each cancer type and all *K* SNPs under an unconstrained *K* degree-of-freedom model, and for the nested 1 degree-of-freedom risk score model, respectively. Then the goodness-of-fit test statistic is:

$$Q_{rs} = X_k^2 - X_{rs}^2 \qquad (5)$$

in which Q_{rs} is $\chi^2_{(k-1)}$ distributed under the null hypothesis that all *K* SNPs are associated with cancer risk with true association sizes proportional to the associations with telomere length. For each analysis in which the goodness-of-fit test null hypothesis was rejected (P<0.05), we removed the SNP that resulted in the greatest reduction of the Q_{rs} test statistic, and repeated the goodness-of-fit test. If still P<0.05, we repeated the exclusion procedure until P>0.05. The association estimates obtained from the subsequent analysis using the remaining SNPs that pass the goodness-of-fit test is referred to as the "goodness-of-fit based" estimates and displayed in Supplementary Table 2.1. P-values for the goodness-of-fit test before and after exclusion of SNPs, and which SNPs were excluded for each cancer analysis for the "goodness-of-fit based" analysis are shown in Supplementary Table 2.2. An example for this procedure is as follows: for lung adenocarcinoma, the inclusion of all 9 SNPs results in an OR of 2.87 (P = 6.3x10^{-15}), but the goodness-of-fit test p-value of 9.0x10^{-6} indicates the presence of heterogeneity with at least one of the SNPs used in the SNP score. The *TERT* SNP was identified to be the SNP contributing the most to this heterogeneity (by stepwise exclusion), and once removed, resulted in a SNP score with a goodness-of-fit p-value of 0.09, indicating a lack of substantial heterogeneity in ratio of SNP-cancer to SNP-telomere associations across SNPs. After exclusion of the *TERT* SNP, we still observe a significant association in the same direction, albeit attenuated (OR=2.00, P = 6.6x10^{-6}).

We performed power analyses to estimate the minimum detectable magnitude of association for each cancer type given the sample sizes available in the GAME-ON study, in

terms of OR per 1 kb increase in TL. This was done using a web-based application (http://glimmer.rstudio.com/kn3in/mRnd/) (69), specifying 80% power, 0.05 type-I error rate, and assuming the variance in TL explained by the nine SNPs is R^2=0.01 or R^2=0.02, respectively. Because the web application calculates the detectable OR of cancer risk per one standard deviation of telomere length, which is roughly equivalent to 500 bp (36, 37), we exponentiated this OR to the power of 0.5 in order to obtain the detectable OR per 1000 bp increase TL. (Supplementary Table 2.3).

ACKNOWLEDGMENTS

Funding

This work was supported by the Genetic Associations and Mechanisms in Oncology Network, GAME-ON: Discovery, Biology, and Risk of Inherited Variants in Breast Cancer, DRIVE, PI: D. Hunter (grant number U19 CA148065); Colorectal Transdisciplinary Study, CORECT, PI: S. Gruber (grant number U19 CA148107); Transdisciplinary Research in Cancer of the Lung, TRICL, PI: C. Amos (grant number U19 CA148127); Follow-up of ovarian cancer genetic association and interaction studies, FOCI, PI: T. Sellers (grant number U19 CA148112); Elucidating Loci Involved in Prostate Cancer Susceptibility, ELLIPSE, PI: B. Henderson (grant number U19 CA148537); Genetics and Epidemiology of Colorectal Cancer Consortium, GECCO: National Cancer Institute, National Institutes of Health, U.S. Department of Health and Human Services (grant numbers U01 CA137088, R01 CA059045); National Institute of Health, National Institute on Aging (T32AG000243; P30AG012857) to CZ; Cancer Research Foundation Young Investigator Award to BLP; The Wellcome Trust (grant number 100114) to SB; The NHMRC Senior Principal Research Fellow to GCT

REFERENCES

1. d'Adda di Fagagna F. Living on a break: cellular senescence as a DNA-damage response. Nat Rev Cancer. 2008;8:512–22.

2. Shawi M, Autexier C. Telomerase, senescence and ageing. Mech Ageing Dev. 2008;129:3–10.

3. Stewart SA, Weinberg RA. Telomeres: Cancer to Human Aging. Annu Rev Cell Dev Biol. 2006;22:531–57.

4. O'Sullivan RJ, Karlseder J. Telomeres: protecting chromosomes against genome instability. Nat Rev Mol Cell Biol. 2010;11:171–81.

5. Willeit P, Willeit J, Mayr A, Weger S, Oberhollenzer F, Brandstätter A, et al. Telomere length and risk of incident cancer and cancer mortality. J Am Med Assoc. 2010;304:69–75.

6. Ma H, Zhou Z, Wei S, Liu Z, Pooley KA, Dunning AM, et al. Shortened Telomere Length Is Associated with Increased Risk of Cancer: A Meta-Analysis. PLoS ONE. 2011;6:e20466.

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

8. Wentzensen IM, Mirabello L, Pfeiffer RM, Savage SA. The Association of Telomere Length and Cancer: a Meta-analysis. Cancer Epidemiol Biomarkers Prev. 2011;20:1238–50.

9. Qin, Q, Sun, J, Yin, J, Liu, L, Chen, J, Zhang, Y, et al. Telomere Length in Peripheral Blood Leukocytes Is Associated with Risk of Colorectal Cancer in Chinese Population. PLoS ONE. 2014;9:e88135.

10. Shen J, Gammon MD, Terry MB, Wang Q, Bradshaw P, Teitelbaum SL, et al. Telomere length, oxidative damage, antioxidants and breast cancer risk. Int J Cancer. 2009;124:1637–43.

11. Qu S, Wen W, Shu X-O, Chow W-H, Xiang Y-B, Wu J, et al. Association of Leukocyte Telomere Length With Breast Cancer Risk: Nested Case-Control Findings From the Shanghai Women's Health Study. Am J Epidemiol. 2013;177:617–24.

12. Weischer M, Nordestgaard BG, Cawthon RM, Freiberg JJ, Tybjærg-Hansen A, Bojesen SE. Short Telomere Length, Cancer Survival, and Cancer Risk in 47102 Individuals. J Natl Cancer Inst. 2013;105:459–68.

13. Lan Q, Cawthon R, Gao Y, Hu W, Hosgood HD III, 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:e59230.

14. Pooley KA, Sandhu MS, Tyrer J, Shah M, Driver KE, Luben RN, et al. Telomere Length in Prospective and Retrospective Cancer Case-Control Studies. Cancer Res. 2010;70:3170–6.

15. Zee RYL, Castonguay AJ, Barton NS, Buring JE. Mean Telomere Length and Risk of Incident Colorectal Carcinoma: A Prospective, Nested Case-Control Approach. Cancer Epidemiol Biomarkers Prev. 2009;18:2280–2.

16. Cui Y, Cai Q, Qu S, Chow W-H, Wen W, Xiang Y-B, et al. Association of Leukocyte Telomere Length with Colorectal Cancer Risk: Nested Case–Control Findings from the Shanghai Women's Health Study. Cancer Epidemiol Biomarkers Prev. 2012;21:1807–13.

17. Lee J-J, Nam C-E, Cho S-H, Park K-S, Chung I-J, Kim H-J. Telomere length shortening in non-Hodgkin's lymphoma patients undergoing chemotherapy. Ann Hematol. 2003;82:492–5.

18. Schroder CP, Wisman GBA, Jong S de, Graaf WTA van der, Ruiters MHJ, Mulder NH, et al. Telomere length in breast cancer patients before and after chemotherapy with or without stem cell transplantation. Br J Cancer. 2001;84:1348–53.

19. 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;45:422–7.

20. Mangino M, Hwang S-J, 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. Hum Mol Genet. 2012;21:5385–94.

21. 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. Hum Mol Genet. 2013;22:5056–64.

22. 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;45:371–84.

23. Peters U, Hutter C, Hsu L, Schumacher F, Conti D, Carlson C, et al. Meta-analysis of new genome-wide association studies of colorectal cancer risk. Hum Genet. 2012;131:217–34.

24. 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;41:221–7.

25. Didelez V, Sheehan N. Mendelian randomization as an instrumental variable approach to causal inference. Stat Methods Med Res. 2007;16:309–30.

26. Hung RJ, Ulrich CM, Goode EL, Brhane Y, Muir K, Chan AT, et al. Cross Cancer Genomic Investigation of Inflammation Pathway for Five Common Cancers: Lung, Ovary, Prostate, Breast and Colorectal Cancer. Journal of the National Cancer Institute. 2015;

27. Dastani Z, Hivert M-F, Timpson N, Perry JRB, Yuan X, Scott RA, et al. Novel Loci for Adiponectin Levels and Their Influence on Type 2 Diabetes and Metabolic Traits: A Multi-Ethnic Meta-Analysis of 45,891 Individuals. PLoS Genet. 2012;8:e1002607.

28. Burgess S, Butterworth A, Thompson SG. Mendelian Randomization Analysis With Multiple Genetic Variants Using Summarized Data. Genet Epidemiol. 2013;37:658–65.

29. Johnson T. Efficient Calculation for Multi-SNP Genetic Risk Scores. San Francisco, CA; 2012.

30. Burgess S, Scott R, Timpson N, Davey Smith G, Thompson S. Using published data in Mendelian randomization: a blueprint for efficient identification of causal risk factors. Eur J Epidemiol. 2015;1–10.

31. 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;95:1211–8.

32. Hosgood III HD, Cawthon R, He X, Chanock S, Lan Q. Genetic variation in telomere maintenance genes, telomere length, and lung cancer susceptibility. Lung Cancer. 2009;66:157–61.

33. 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 Res. 2014;74:2476–86.

34. Shen M, Cawthon R, Rothman N, Weinstein SJ, Virtamo J, Hosgood III HD, et al. A prospective study of telomere length measured by monochrome multiplex quantitative PCR and risk of lung cancer. Lung Cancer. 2011;73:133–7.

35. Seow WJ, Cawthon RM, Purdue MP, Hu W, Gao Y-T, Huang W-Y, et al. Telomere Length in White Blood Cell DNA and Lung Cancer: A Pooled Analysis of Three Prospective Cohorts. Cancer Res. 2014;74:4090–8.

36. Shaffer JA, Epel E, Kang MS, Ye S, Schwartz JE, Davidson KW, et al. Depressive Symptoms Are Not Associated with Leukocyte Telomere Length: Findings from the Nova Scotia Health Survey (NSHS95), a Population-Based Study. PLoS ONE. 2012;7:e48318.

37. Puterman E, Lin J, Krauss J, Blackburn EH, Epel ES. Determinants of telomere attrition over 1 year in healthy older women: stress and health behaviors matter. Mol Psychiatry. 2015;20:529–35.

38. Valdes A, Andrew T, Gardner J, Kimura M, Oelsner E, Cherkas L, et al. Obesity, cigarette smoking, and telomere length in women. The Lancet. 20;366:662–4.

39. Timofeeva MN, Hung RJ, Rafnar T, Christiani DC, Field JK, Bickeböller H, et al. Influence of common genetic variation on lung cancer risk: meta-analysis of 14 900 cases and 29 485 controls. Hum Mol Genet. 2012;21:4980–95.

40. Daraselia N, Wang Y, Budoff A, Lituev A, Potapova O, Vansant G, et al. Molecular signature and pathway analysis of human primary squamous and adenocarcinoma lung cancers. Am J Cancer Res. 2012;2:98–103.

41. Sekido Y, Fong KM, Minna JD. Molecular Genetics of Lung Cancer. Annu Rev Med. 2003;54:73–87.

42. Petersen I, Bujard M, Petersen S, Wolf G, Goeze A, Schwendel A, et al. Patterns of Chromosomal Imbalances in Adenocarcinoma and Squamous Cell Carcinoma of the Lung. Cancer Res. 1997;57:2331–5.

43. Machiela MJ, Hsiung CA, Shu X-O, Seow WJ, Wang Z, Matsuo K, et al. Genetic variants associated with longer telomere length are associated with increased lung cancer risk among never-smoking women in Asia: a report from the female lung cancer consortium in Asia. Int J Cancer. 2015;137:311–9.

44. Fagagna F d'Adda di, Reaper PM, Clay-Farrace L, Fiegler H, Carr P, von Zglinicki T, et al. A DNA damage checkpoint response in telomere-initiated senescence. Nature. 2003;426:194–8.

45. Bull CF, Mayrhofer G, O'Callaghan NJ, Au AY, Pickett HA, Low GKM, 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 Prev Res Phila Pa. 2014;7:128–38.

46. Halaschek-Wiener J, Vulto I, Fornika D, Collins J, Connors JM, Le ND, et al. Reduced telomere length variation in healthy oldest old. Mech Ageing Dev. 2008;129:638–41.

47. Iles MM, Bishop DT, Taylor JC, Hayward NK, Brossard M, Cust AE, et al. The effect on melanoma risk of genes previously associated with telomere length. J Natl Cancer Inst. 2014;106:dju267.

48. Bataille V, Kato BS, Falchi M, Gardner J, Kimura M, Lens M, et al. Nevus Size and Number Are Associated with Telomere Length and Represent Potential Markers of a Decreased Senescence In vivo. Cancer Epidemiol Biomarkers Prev. 2007;16:1499–502.

49. Aviv A, Chen W, Gardner JP, Kimura M, Brimacombe M, Cao X, et al. Leukocyte Telomere Dynamics: Longitudinal Findings Among Young Adults in the Bogalusa Heart Study. Am J Epidemiol. 2009;169:323–9.

50. Lubin JH, Blot WJ. Assessment of lung cancer risk factors by histologic category. J Natl Cancer Inst. 1984;73:383–9.

51. Hurwitz LM, Heaphy CM, Joshu CE, Isaacs WB, Konishi Y, De Marzo AM, et al. Telomere length as a risk factor for hereditary prostate cancer. The Prostate. 2014;74:359–64.

52. Mirabello L, Huang W-Y, Wong JYY, Chatterjee N, Reding D, David Crawford E, et al. The association between leukocyte telomere length and cigarette smoking, dietary and physical variables, and risk of prostate cancer. Aging Cell. 2009;8:405–13.

53. Jones AM, Beggs AD, Carvajal-Carmona L, Farrington S, Tenesa A, Walker M, et al. TERC polymorphisms are associated both with susceptibility to colorectal cancer and with longer telomeres. Gut. 2012;61:248–54.

54. Peters U, Jiao S, Schumacher FR, Hutter CM, Aragaki AK, Baron JA, et al. Identification of Genetic Susceptibility Loci for Colorectal Tumors in a Genome-Wide Meta-analysis. Gastroenterology. 2013;144:799–807.e24.

55. Martinez-Delgado B, Yanowsky K, Inglada-Perez L, de la Hoya M, Caldes T, Vega A, et al. Shorter telomere length is associated with increased ovarian cancer risk in both familial and sporadic cases. J Med Genet. 2012;49:341–4.

56. Mirabello L, Garcia-Closas M, Cawthon R, Lissowska J, Brinton L, Pepłońska B, et al. Leukocyte telomere length in a population-based case–control study of ovarian cancer: a pilot study. Cancer Causes Control. 2010;21:77–82.

57. Terry KL, Tworoger SS, Vitonis AF, Wong J, Titus-Ernstoff L, De Vivo I, et al. Telomere Length and Genetic Variation in Telomere Maintenance Genes in Relation to Ovarian Cancer Risk. Cancer Epidemiol Biomarkers Prev. 2012;21:504–12.

58. Kyo S, Takakura M, Kanaya T, Zhuo W, Fujimoto K, Nishio Y, et al. Estrogen Activates Telomerase. Cancer Res. 1999;59:5917–21.

59. Lin J, Kroenke CH, Epel E, Kenna HA, Wolkowitz OM, Blackburn E, et al. Greater endogenous estrogen exposure is associated with longer telomeres in postmenopausal women at risk for cognitive decline. Window Oppor Menopause Estrogens Brain. 2011;1379:224–31.

60. Martínez P, Blasco MA. Telomeric and extra-telomeric roles for telomerase and the telomere-binding proteins. Nat Rev Cancer. 2011;11:161–76.

61. Anderson BH, Kasher PR, Mayer J, Szynkiewicz M, Jenkinson EM, Bhaskar SS, et al. Mutations in CTC1, encoding conserved telomere maintenance component 1, cause Coats plus. Nat Genet. 2012;44:338–42.

62. Pierce BL, Ahsan H, Vanderweele TJ. Power and instrument strength requirements for Mendelian randomization studies using multiple genetic variants. Int J Epidemiol. 2011;40:740–52.

63. Saferali A, Lee J, Sin DD, Rouhani FN, Brantly ML, Sandford AJ. Longer Telomere Length in COPD Patients with α1-Antitrypsin Deficiency Independent of Lung Function. PLoS ONE. 2014;9:e95600.

64. Alder JK, Chen JJ-L, Lancaster L, Danoff S, Su S, Cogan JD, et al. Short telomeres are a risk factor for idiopathic pulmonary fibrosis. Proc Natl Acad Sci. 2008;105:13051–6.

65. 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.

66. Aviv A, Hunt SC, Lin J, Cao X, Kimura M, Blackburn E. Impartial comparative analysis of measurement of leukocyte telomere length/DNA content by Southern blots and qPCR. Nucleic Acids Res. 2011;39:e134–e134.

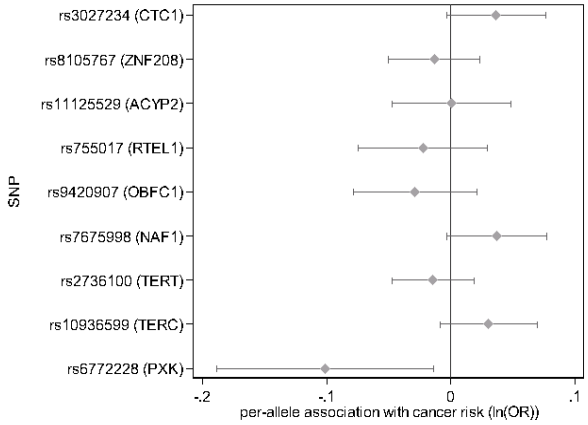
67. Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, de Bakker PIW. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. Bioinformatics. 2008;24:2938–9.

68. Eeles RA, Olama AAA, Benlloch S, Saunders EJ, Leongamornlert DA, Tymrakiewicz M, et al. Identification of 23 new prostate cancer susceptibility loci using the iCOGS custom genotyping array. Nat Genet. 2013;45:385–91, 391-2.

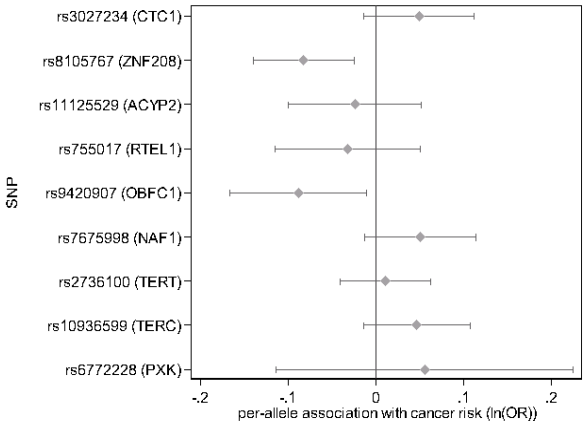
69. Brion M-JA, Shakhbazov K, Visscher PM. Calculating statistical power in Mendelian randomization studies. Int J Epidemiol. 2013;42:1497–501.

SUPPLEMENTARY MATERIALS

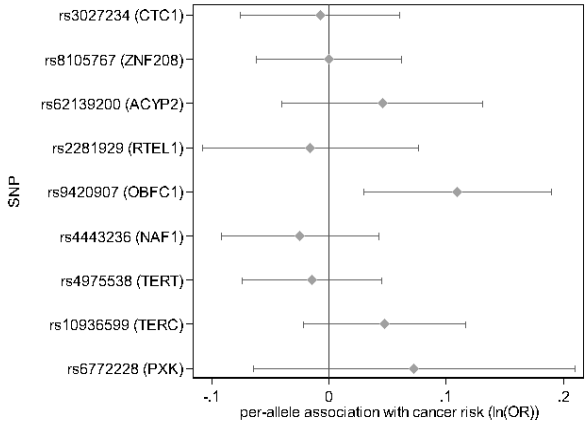
Supplementary Figure 2.1. Forest plots of association estimates (with horizontal bars indicating 95% confidence interval) for the "long telomere" allele of each SNP with cancer risk. SNPs are ordered by increasing magnitude of association with telomere length



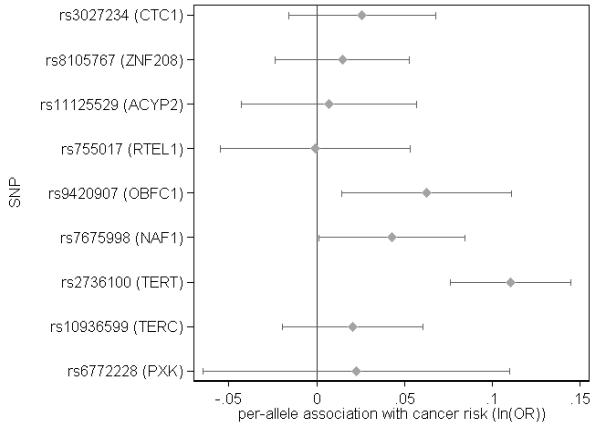
Breast cancer – all subtypes



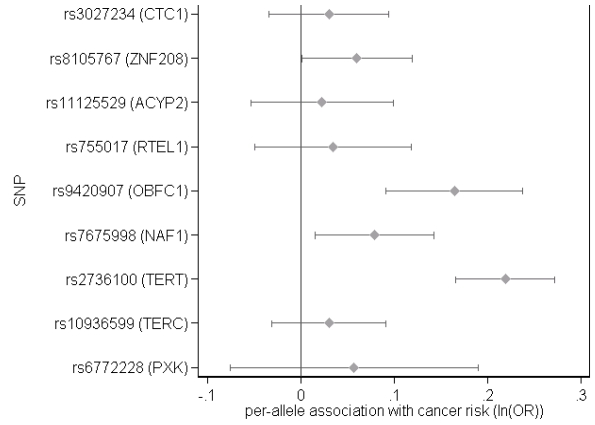
Breast cancer – ER-negative



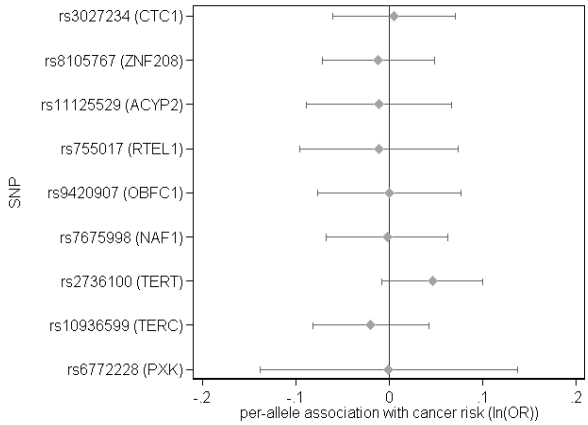
Colorectal cancer – all subtypes



Lung cancer – all subtypes

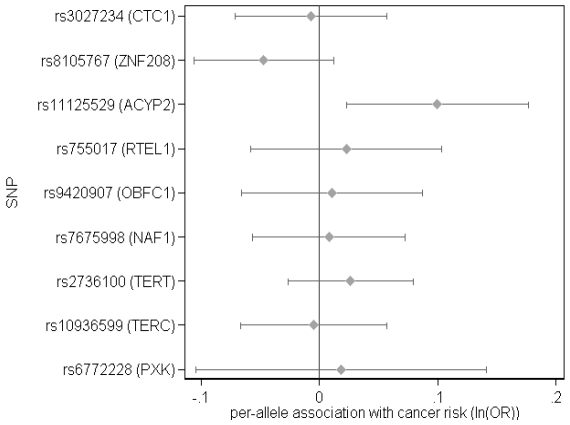


Lung cancer - adenocarcinoma

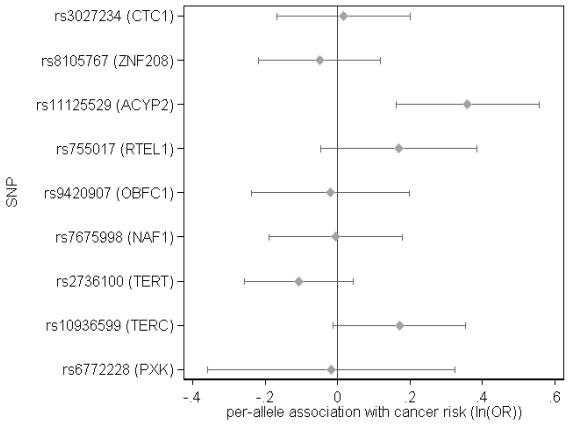


Lung cancer - squamous

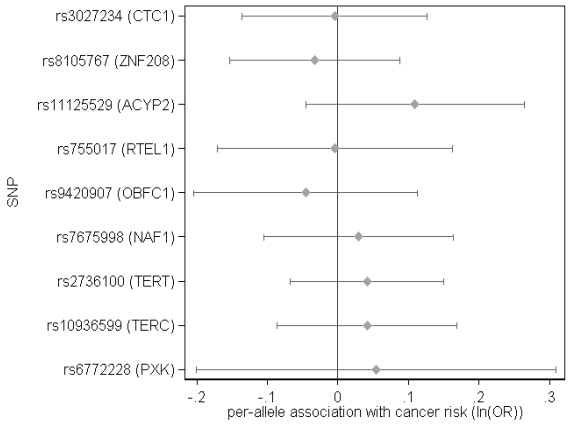
Supplementary Figure 2.1. Continued



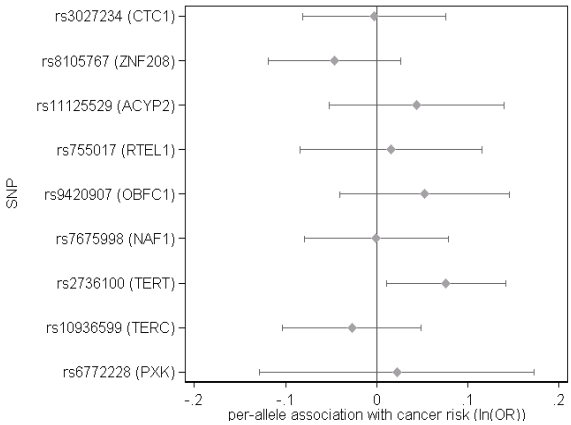
Ovarian cancer – all subtypes



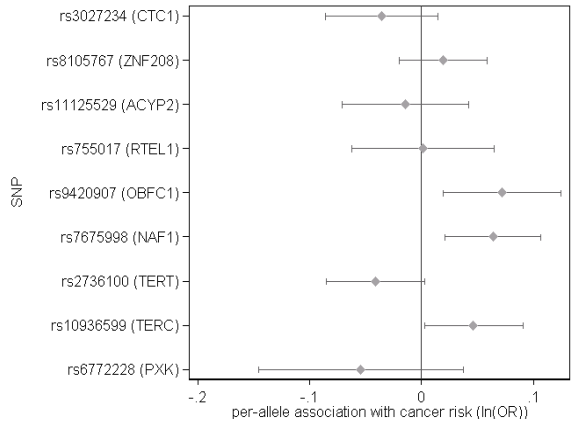
Ovarian cancer – clear cell



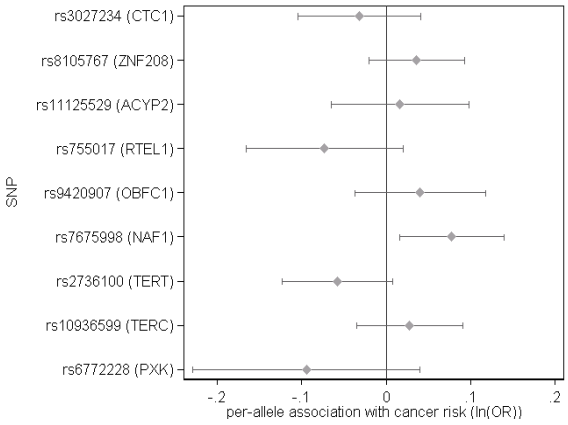
Ovarian cancer – endometrioid



Ovarian cancer – serous

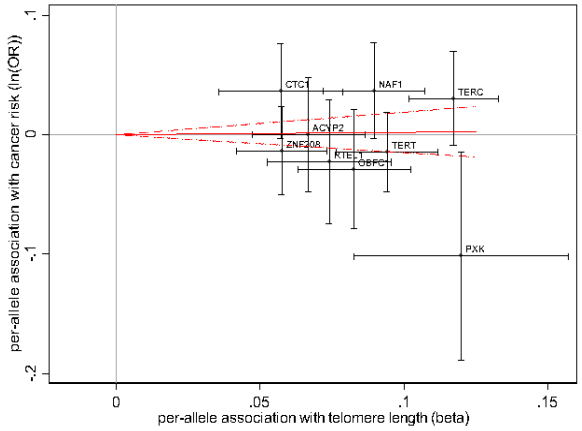


Prostate cancer – all subtypes

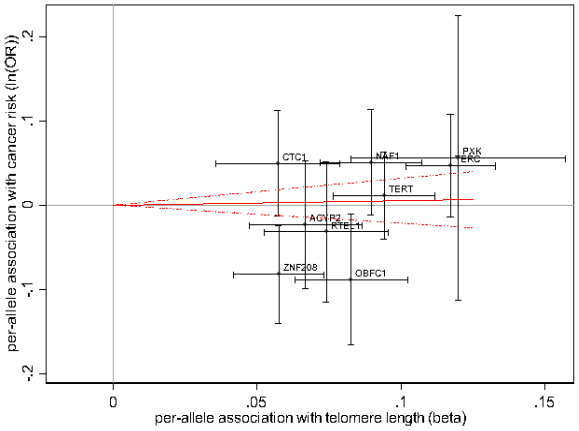


Prostate cancer - aggressive

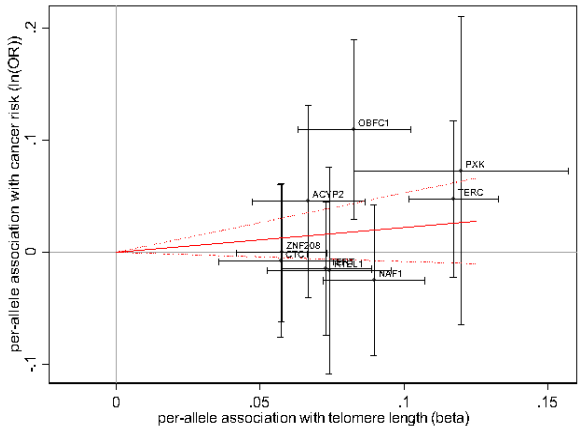
Supplementary Figure 2.2. Scatter plots showing the per-allele association with cancer risk plotted against the per-allele association with kb of telomere length (with vertical and horizontal black lines showing 95% confidence interval for each SNP). The scatter plot is overlaid with the Mendelian randomization estimate (slope of red solid line with dotted lines showing 95% confidence interval) of the effect of TL on cancer risk



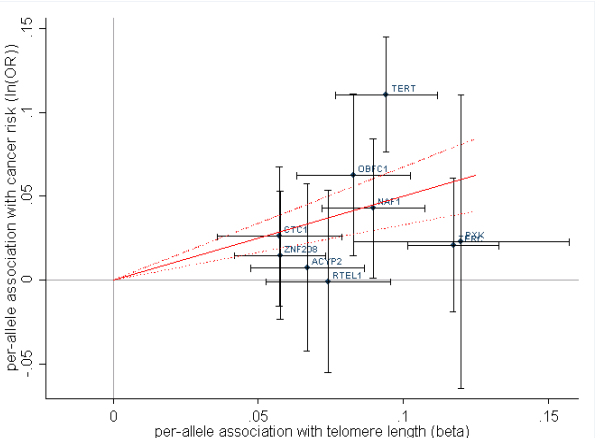
Breast cancer – all subtypes



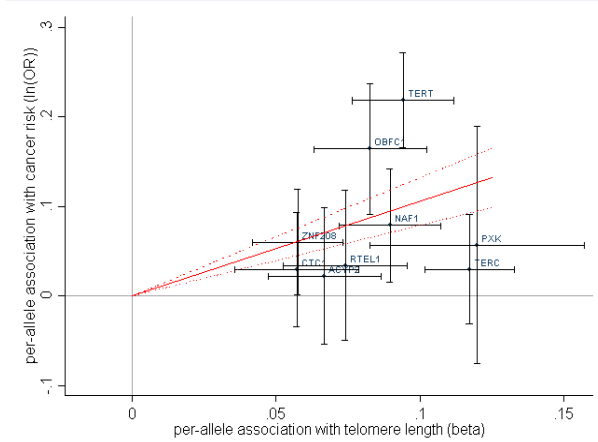
Breast cancer – ER-negative



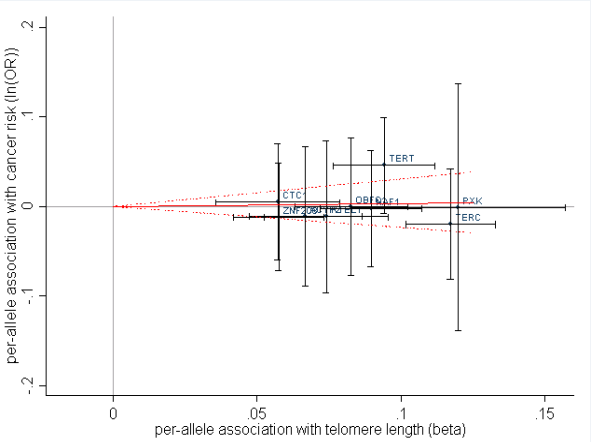
Colorectal cancer – all subtypes



Lung cancer – all subtypes

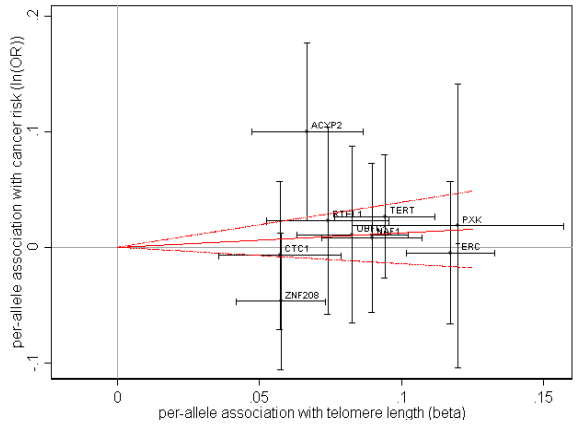


Lung cancer - adenocarcinoma

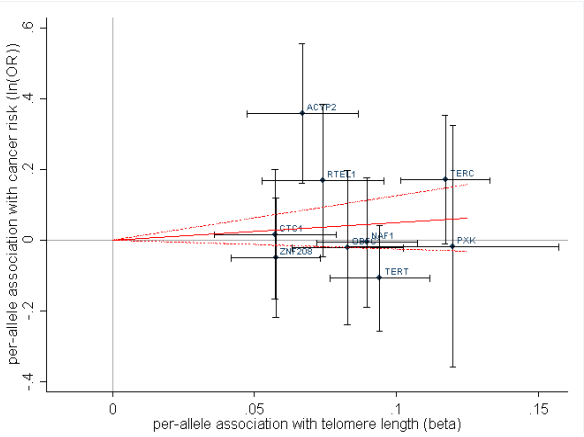


Lung cancer - squamous

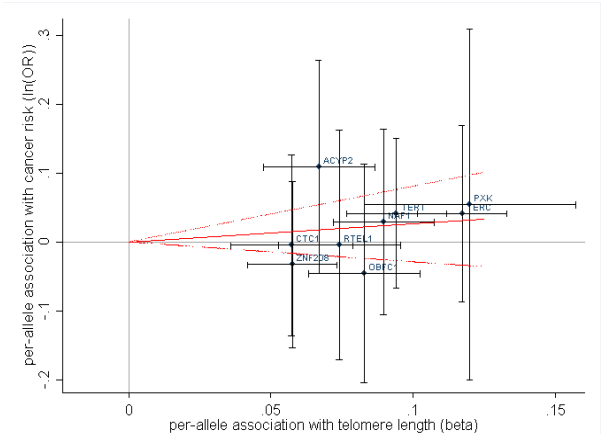
Supplementary Figure 2.2. Continued



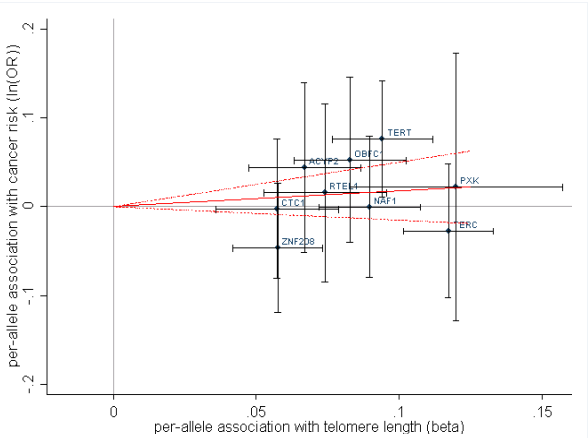
Ovarian cancer – all subtypes



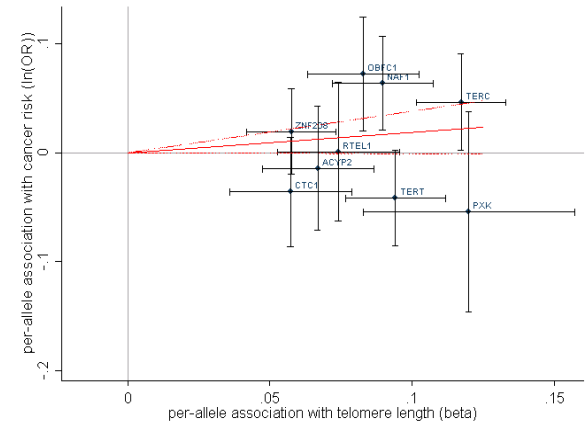
Ovarian cancer – clear cell



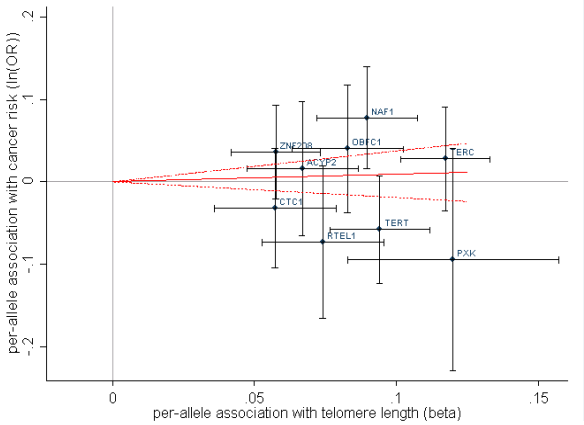
Ovarian cancer – endometrioid



Ovarian cancer – serous

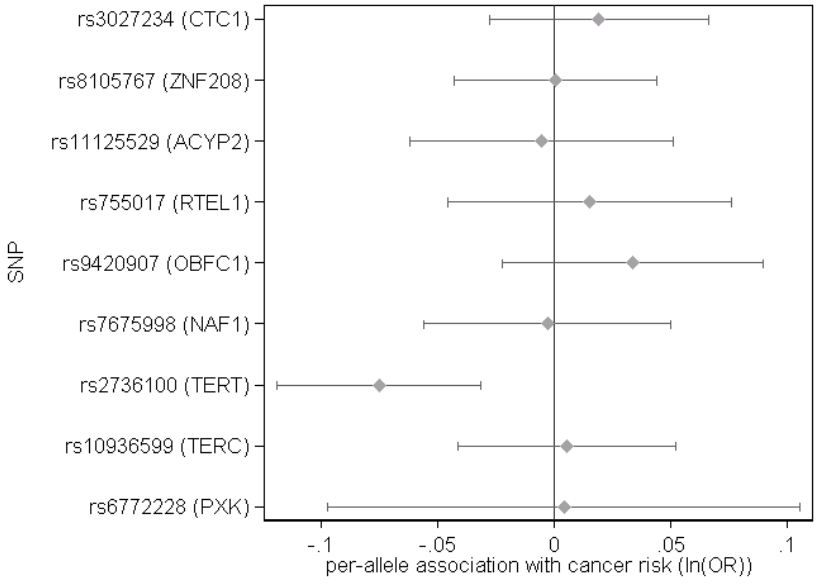


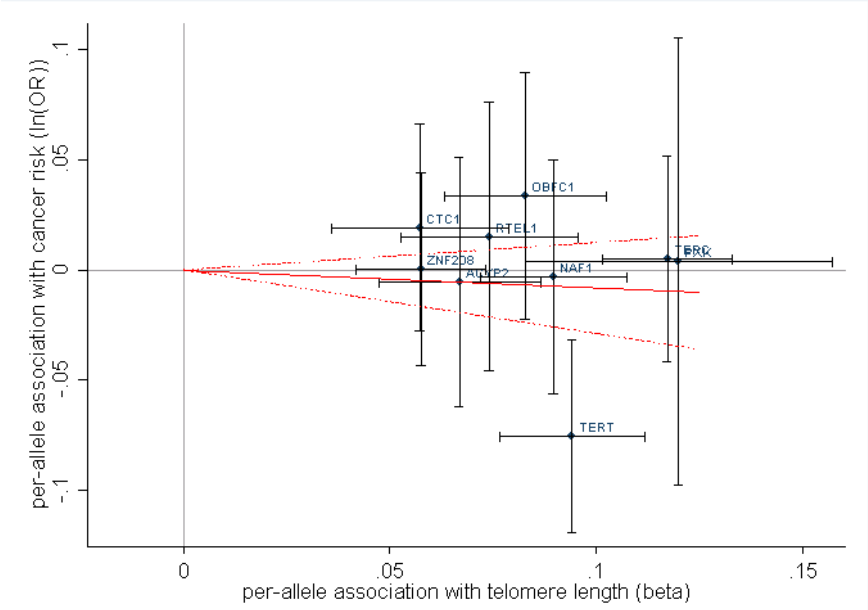
Prostate cancer – all subtypes



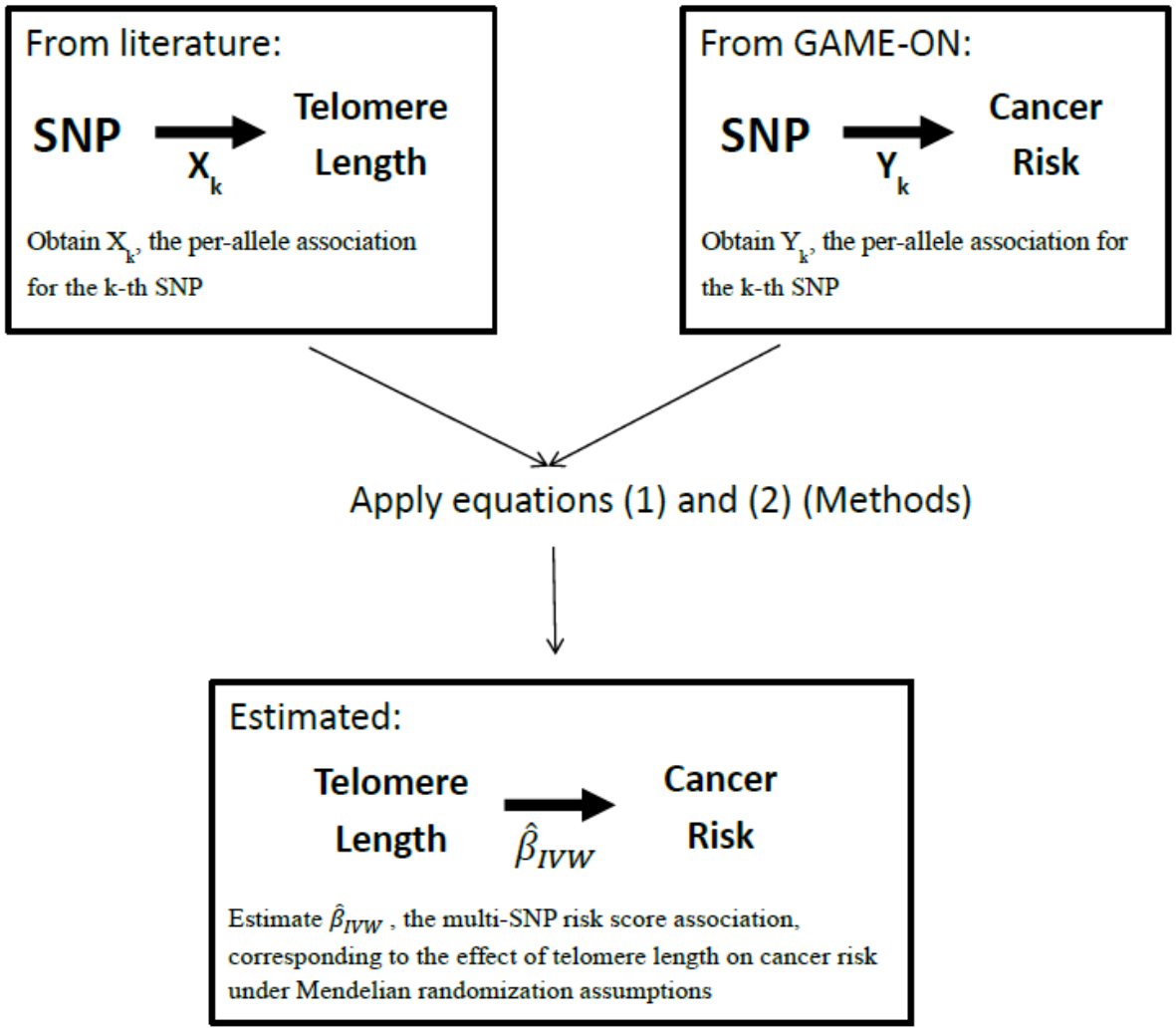
Prostate cancer - aggressive

Supplementary Figure 2.3. Forest plot (top) and scatter plot (bottom) of associations between telomere length-associated SNPs and risk for colorectal cancer. Forest plots show association estimates (with horizontal bars indicating 95% confidence interval) for the "long telomere" allele of each SNP with cancer risk. SNPs are ordered by increasing magnitude of association with telomere length. Scatter plots show the per-allele association with cancer risk plotted against the per-allele association with kb of telomere length (with vertical and horizontal black lines showing 95% confidence interval for each SNP). The scatter plot is overlaid with the Mendelian randomization estimate (slope of red solid line with dotted lines showing 95% confidence interval) of the effect of TL on cancer risk Data is from 10,314 independent cases and 12,857 controls from the Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO)





Supplementary Figure 2.4. Schematic of the Mendelian randomization estimation method using data from two different sources



Supplementary Table 2.1. Odds ratios of cancer risk per 1000 base pair increase in telomere length according to a multi-SNP telomere length score analysis of strict[a], goodness-of-fit (GoF) based[b], and strict-goodness-of-fit based[c] list of SNPs.

	Strict			GoF-based			Strict + GoF-based		
Cancer Type	Estimate	95% CI	P_{fit}	Estimate	95% CI	P_{fit}	Estimate	95% CI	P_{fit}
Breast									
All	1.07	0.90, 1.28	0.43	1.07	0.90, 1.28	0.43	1.07	0.90, 1.28	0.43
ER-negative	1.04	0.79, 1.36	0.79	1.18	0.89, 1.55	0.24	1.16	0.88, 1.54	0.29
Colorectal	1.21	0.89, 1.66	0.23	1.25	0.92, 1.69	0.15	1.21	0.89, 1.66	0.23
Lung									
All	1.68	1.41, 2.01	$1.0x10^{-8}$	1.36	1.12, 1.65	$2.3x10^{-3}$	1.37	1.12, 1.68	$2.5x10^{-3}$
Adeno-carcinoma	2.93	2.26, 3.80	$5.9x10^{-16}$	2.00	1.48, 2.70	$6.6x10^{-6}$	2.06	1.53, 2.78	$1.9x10^{-6}$
Squamous	0.98	0.74, 1.29	0.87	1.04	0.79, 1.36	0.79	0.98	0.74, 1.29	0.87
Ovarian									
All	1.13	0.86, 1.48	0.39	1.13	0.87, 1.47	0.37	1.13	0.86, 1.48	0.39
Clear-cell	1.73	0.79, 3.79	0.17	1.18	0.54, 2.57	0.68	1.21	0.54, 2.72	0.65
Endometrioid	1.28	0.73, 2.25	0.40	1.30	0.75, 2.24	0.35	1.28	0.73, 2.25	0.40
Serous	1.19	0.85, 1.67	0.32	1.19	0.86, 1.65	0.30	1.19	0.85, 1.67	0.32
Prostate									
All	1.26	1.03, 1.54	0.02	1.46	1.17, 1.82	$7.9x10^{-4}$	1.47	1.17, 1.83	$7.5x10^{-4}$
Aggressive	1.16	0.87, 1.56	0.31	0.94	0.69, 1.28	0.68	1.37	0.99, 1.89	0.05

[a]Strict set includes only high-confidence SNPs.
[b]GoF-based set includes only non-pleiotropic SNPs based on goodness-of-fit test
[c]Strict-GoF-based set includes only high-confidence, non-pleiotropic SNPs

Supplementary Table 2.2. P-values (P_{fit}) for the goodness-of-fit test before and after exclusion of SNPs, and the SNPs excluded for each cancer analysis for the "goodness-of-fit based[a]" analysis.

Cancer Type	P-value of goodness-of-fit test statistic including all SNPs	SNPs excluded in stepwise procedure	P-value of the goodness-of-fit test statistic after SNP exclusion
Breast			
All	0.03	rs3027234 (PXK)	0.12
ER-negative	$6.5x10^{-3}$	rs8105767 (ZNF208)	0.09
Colorectal	0.22	-	-
Lung			
All	$5.8x10^{-3}$	rs2736100 (TERT)	0.70
Adenocarcinoma	$9.0x10^{-6}$	rs2736100 (TERT)	0.09
Squamous	0.90	-	-
Ovarian			
All	0.29	-	-
Clear-cell	0.01	rs11125529 (ACYP2)	0.34
Endometrioid	0.94	-	-
Serous	0.40	-	-
Prostate			
All	$2.0x10^{-3}$	rs2736100, rs3027234 (TERT, CTC1)	0.08
Aggressive	0.03	rs7675998 (NAF1)	0.13

[a]The goodness-of-fit test described by Johnson (http://cran.r-project.org/web/packages/gtx/index.html) was applied to the nine SNPs used in the multi-SNP score for each cancer risk association analysis. A chi-squared test statistic was calculated under the null hypothesis that all nine SNPs are associated with cancer risk with true effect sizes proportional to the effects on telomere length. For each analysis in which the goodness-of-fit test null hypothesis was rejected ($p<0.05$), we removed the SNP that resulted in the greatest reduction of the test statistic. For all but one cancer subtype analysis, the exclusion of one SNP removed evidence of heterogeneity ($p>0.05$). For the overall prostate cancer analysis, a second SNP was excluded based on the above described method, resulting in no remaining evidence of heterogeneity. All remaining SNPs that passed the "goodness-of-fit" test was evaluated as multi-SNP scores in association with cancer risk.

Supplementary Table 2.3. Approximate odds ratio (OR) detectable per 2 standard deviations of telomere length (1000 bp) given sample size, case proportion, and 80% power with 0.05 type-I error rate assuming the variance in TL explained by the nine SNPs is R^2=0.01 or R^2=0.02, respectively.

Cancer Type	Sample size	Proportion cases	OR (R^2=0.01)	OR (R^2=0.02)
Breast				
All	33,832	0.47	0.86/1.16	0.90/1.11
ER-negative	18,067	0.27	0.81/1.24	0.85/1.17
Colorectal	9,931	0.51	0.76/1.32	0.82/1.22
Colorectal (GECCO)	21,528	0.44	0.83/1.20	0.88/1.14
Lung[a]				
All	28,998	0.42	0.85/1.17	0.89/1.12
Adenocarcinoma	19,589	0.19	0.80/1.24	0.85/1.17
Squamous	19,437	0.18	0.80/1.25	0.85/1.18
Ovarian[a]				
All	13,492	0.55	0.79/1.4	0.84/1.17
Clear-cell	9,479	0.10	0.63/1.62	0.70/1.44
Endometrioid	9,836	0.18	0.68/1.52	0.74/1.35
Serous	11,696	0.44	0.77/1.41	0.82/1.22
Prostate				
All	26,884	0.53	0.84/1.19	0.88/1.13
Aggressive	17,174	0.26	0.80/1.24	0.85/1.17

Supplementary Table 2.4. Proxy SNPs of genetic variants associated with long relative telomere length from literature used for colorectal cancer association analyses

Chromosome	Locus	Original SNP	Proxy SNP	Distance	r^2	Proxy Long Allele
2	ACYP2	rs11125529	rs62139200	359	1.000	T
4	NAF1	rs7675998	rs4443236	3316	0.953	T
5	TERT	rs7726159	rs4975538	1489	0.930	C
20	RTEL1	rs755017	rs2281929	458	1.000	C

Supplementary Table 2.5. Proxy SNPs of genetic variants associated with long relative telomere length from literature used for age and sex stratified overall lung cancer association analyses

Chromosome	Locus	Original SNP	Proxy SNP	Distance	r^2	Proxy Long Allele
4	NAF1	rs7675998	rs11100479	5081	0.953	T
19	ZNF208	rs8105767	rs7257051	103851	0.853	G
20	RTEL1	rs755017	rs2281929	458	1.000	C
2	ACYP2	rs11125529	rs11890390	9816	0.932	T
17	CTC1	rs3027234	rs8075078	50004	0.862	G

Funding and Acknowledgements:

BCFR:

The Breast Cancer Family Registry (BCFR) is supported by grant UM1 CA164920 from the USA National Cancer Institute. The content of this manuscript does not necessarily reflect the views or policies of the National Cancer Institute or any of the collaborating centers in the Breast Cancer Family Registry (BCFR), nor does mention of trade names, commercial products, or organizations imply endorsement by the USA Government or the BCFR.

GECCO:

The authors would like to thank all those at the GECCO Coordinating Center for helping bring together the data and people that made this project possible. The authors acknowledge Dave Duggan and team members at TGEN (Translational Genomics Research Institute), the Broad Institute, and the Génome Québec Innovation Center for genotyping DNA samples of cases and controls, and for scientific input for GECCO.

ASTERISK: a Hospital Clinical Research Program (PHRC) and supported by the Regional Council of Pays de la Loire, the Groupement des Entreprises Françaises dans la Lutte contre le Cancer (GEFLUC), the Association Anne de Bretagne Génétique and the Ligue Régionale Contre le Cancer (LRCC). We are very grateful to Dr. Bruno Buecher without whom this project would not have existed. We also thank all those who agreed to participate in this study, including the patients and the healthy control persons, as well as all the physicians, technicians and students.

COLO2&3: National Institutes of Health (R01 CA60987).

DACHS: German Research Council (Deutsche Forschungsgemeinschaft, BR 1704/6-1, BR 1704/6-3, BR 1704/6-4 and CH 117/1-1), and the German Federal Ministry of Education and Research (01KH0404 and 01ER0814). We thank all participants and cooperating clinicians, and Ute Handte-Daub, Renate Hettler-Jensen, Utz Benscheid, Muhabbet Celik and Ursula Eilber for excellent technical assistance.

DALS: National Institutes of Health (R01 CA48998 to M.L.S);

HPFS is supported by the National Institutes of Health (P01 CA 055075, UM1 CA167552, R01 137178, R01 CA 151993 and P50 CA 127003), NHS by the National Institutes of Health (R01 CA137178, P01 CA 087969, R01 CA151993 and P50 CA 127003,) and PHS by the National Institutes of Health (R01 CA042182). We would like to acknowledge Patrice Soule and Hardeep Ranu of the Dana Farber Harvard Cancer Center High-Throughput Polymorphism Core who assisted in the genotyping for NHS, HPFS, and PHS under the supervision of Dr. Immaculata Devivo and Dr. David Hunter, Qin (Carolyn) Guo and Lixue Zhu who assisted in programming for NHS and HPFS, and Haiyan Zhang who assisted in programming for the PHS. We would like to thank the participants and staff of the Nurses' Health Study and the Health Professionals Follow-Up Study, for their valuable contributions as well as the following state cancer registries for their help: AL, AZ, AR, CA, CO, CT, DE, FL, GA, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA, WY. The authors assume full responsibility for analyses and interpretation of these data.

MEC: National Institutes of Health (R37 CA54281, P01 CA033619, and R01 CA63464).

OFCCR: National Institutes of Health, through funding allocated to the Ontario Registry for Studies of Familial Colorectal Cancer (U01 CA074783); see CCFR section above. Additional funding toward genetic analyses of OFCCR includes the Ontario Research Fund, the Canadian Institutes of Health Research, and the Ontario Institute for Cancer Research, through generous support from the Ontario Ministry of Research and Innovation.

PLCO: Intramural Research Program of the Division of Cancer Epidemiology and Genetics and supported by contracts from the Division of Cancer Prevention, National Cancer Institute, NIH, DHHS. Additionally, a subset of control samples were genotyped as part of the Cancer Genetic Markers of Susceptibility (CGEMS) Prostate Cancer GWAS (Yeager, M et al. Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. Nat Genet 2007 May;39(5):645-9), Colon CGEMS pancreatic cancer scan (PanScan) (Amundadottir, L et al. Genome-wide association study identifies variants in the ABO locus associated with susceptibility to pancreatic cancer. Nat Genet. 2009 Sep;41(9):986-90, and Petersen, GM et al. A genome-wide association study identifies pancreatic cancer susceptibility loci on chromosomes 13q22.1, 1q32.1 and 5p15.33. Nat Genet. 2010 Mar;42(3):224-8), and the Lung Cancer and Smoking study (Landi MT, et al. A genome-wide association study of lung cancer identifies a region of chromosome 5p15 associated with risk for adenocarcinoma. Am J Hum Genet. 2009 Nov;85(5):679-91). The prostate and PanScan study datasets were accessed with appropriate approval through the dbGaP online resource (http://cgems.cancer.gov/data/) accession numbers

phs000207.v1.p1 and phs000206.v3.p2, respectively, and the lung datasets were accessed from the dbGaP website (http://www.ncbi.nlm.nih.gov/gap) through accession number phs000093.v2.p2. Funding for the Lung Cancer and Smoking study was provided by National Institutes of Health (NIH), Genes, Environment and Health Initiative (GEI) Z01 CP 010200, NIH U01 HG004446, and NIH GEI U01 HG 004438. For the lung study, the GENEVA Coordinating Center provided assistance with genotype cleaning and general study coordination, and the Johns Hopkins University Center for Inherited Disease Research conducted genotyping. The authors thank Drs. Christine Berg and Philip Prorok, Division of Cancer Prevention, National Cancer Institute, the Screening Center investigators and staff or the Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial, Mr. Tom Riley and staff, Information Management Services, Inc., Ms. Barbara O'Brien and staff, Westat, Inc., and Drs. Bill Kopp, Wen Shao, and staff, SAIC-Frederick. Most importantly, we acknowledge the study participants for their contributions to making this study possible. The statements contained herein are solely those of the authors and do not represent or imply concurrence or endorsement by NCI.

PMH: National Institutes of Health (R01 CA076366 to P.A. Newcomb). The authors would like to thank the study participants and staff of the Hormones and Colon Cancer study.

VITAL: National Institutes of Health (K05 CA154337).

WHI: The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through contracts HHSN268201100046C, HHSN268201100001C, HHSN268201100002C,

HHSN268201100003C, HHSN268201100004C, and HHSN271201100004C. The authors thank the WHI investigators and staff for their dedication, and the study participants for making the program possible. A full listing of WHI investigators can be found at: http://www.whi.org/researchers/Documents%20%20Write%20a%20Paper/WHI%20Investigator%20Short%20List.pdf

CHAPTER 3: ASSOCIATION BETWEEN SMOKING AND TELOMERE LENGTH AMONG OLDER ADULTS VARIES BY SEX AND TIMING OF EXPOSURE*

Chenan Zhang, Diane S. Lauderdale, and Brandon L. Pierce

Department of Public Health Sciences, The University of Chicago, Chicago, Illinois (Chenan Zhang, Diane S. Lauderdale, Brandon L. Pierce); Department of Human Genetics, The University of Chicago, Chicago, Illinois (Brandon L. Pierce); and Comprehensive Cancer Center, The University of Chicago, Chicago, Illinois (Brandon L. Pierce)

---

*This is a pre-copyedited, author-produced PDF of an article accepted for publication in American Journal of Epidemiology following peer review. The article is currently in press at the time of publication of this dissertation.

ABSTRACT

Inconsistent associations between smoking and telomere length (TL) have been reported in epidemiologic studies, perhaps due to the time-varying nature of smoking behavior. We estimated the associations between TL (n=5,624, measured by quantitative polymerase chain reaction using saliva DNA) and concurrent and past smoking status, reported biennially for up to 16 years prior to TL measurement, using the Health and Retirement Study (1992-2008). Smoking was associated with reduced TL when using prospective data on smoking status among men and women, but the association was strongly attenuated in cross-sectional analyses among men. This attenuation was largely due to a higher rate of smoking cessation during the study period among males with shorter TL compared to males with longer TL. Short TL was also associated with poorer overall health in men, suggesting that male smokers with short TL were more likely to quit smoking due to poor health. Analyses of years since cessation, smoking duration, and pack-years all support the hypothesis that increased cigarette use shortens TL. Our results provide a potential explanation for inconsistent associations between smoking and TL reported in previous cross-sectional studies. Time-varying associations should be considered in future studies of smoking behavior, TL, aging, and disease risk.

INTRODUCTION

Telomeres are DNA-protein complexes at chromosome ends that protect DNA from damage. Shortening of the DNA component of telomeres occurs with cell division, with average leukocyte telomere length (TL) decreasing approximately 25 to 50 base pairs per year in adults (1). TL has been studied extensively as a biomarker of aging and susceptibility to age-related diseases, including cardiovascular diseases, cancers, and neurocognitive diseases (2–5).

Cigarette smoking results in exposure to free radicals and reactive oxygen species leading to increased oxidative stress and inflammation (6). Both processes have also been linked to TL shortening (7–9), leading to the hypothesis that smoking promotes TL shortening (10). However, epidemiologic reports of smoking and TL have been inconsistent, with some studies finding short TL associated with smoking (11–20), while others find no association (21–26). This inconsistency is also seen in studies of smoking and longitudinal measures of TL attrition, with some studies finding associations between smoking and TL shortening (17,18), while others find no association (19,20). Furthermore, a study of older men in the Helsinki Businessmen Study found that smoking during midlife was associated with short TL measured later in life, but smoking was not associated with TL in cross-sectional analyses conducted in later life (27), suggesting that the timing of smoking behavior assessment may be important.

In this study, we examined the association of TL measured in 2008 with both concurrent smoking and prior smoking reported every two years for up to 16 years prior to TL measurement, among 5,624 participants in the Health and Retirement Study, a population-based study of subjects 50 years and older in the U.S. Since both TL (28–30)) and smoking behavior (31,32) differ by gender, we further examined whether the smoking and TL association varies by sex.

METHODS

Study participants

The Health and Retirement Study (HRS) is a nationally representative longitudinal study of >26,000 subjects aged 50 years and older (33). Sponsored by the National Institute on Aging (U01AG009740) and conducted by the University of Michigan (34), the study collects individual-level information on social, economic and physical health characteristics every two years, beginning in 1992. At Wave 9 (2008), saliva was collected from a subset of individuals to obtain DNA using an Oragene self-collection kit (DNA Genotek, Kanata, Ontario, Canada).

TL measurement

TL was measured using quantitative polymerase chain reaction for subjects who provided a saliva sample (n=5,808) from which DNA was extracted. In brief, TL was measured by comparing telomere sequence copy number (T) to a single-copy gene copy number (S), with the resulting "T/S ratio" proportional to the mean TL of a sample. Assays were run in either duplicate or triplicate, with coefficient of variation of 2.1% between the two procedures. Sample replicates with coefficient of variation greater than 12.5% were reassayed. Control DNA from three cancer cell lines were included on each plate, showing inter-plate assay coefficient of variations 3.5-6.3%. Additional details are available in the 2008 Telomere Length Data Documentation (34).

To reduce skewness, TL was (natural) log transformed. Subjects with log-transformed TL exceeding three standard deviations from the mean (n=66) and subjects younger than age 50 at the time of enrolment (partners of original subjects; n=118) were excluded prior to analyses, resulting in 5,624 remaining subjects. Because the log-transformed TL distribution was still not

normally distributed (skewness=-0.50, kurtosis=6.69), we repeated analyses using a normal-quantile transformation of TL.

Smoking status and covariates

At each wave, data was collected on smoking status (never, former, current) and cigarettes smoked per day for current smokers (<10, 10-<20, ≥20). Also collected at least once during the study for each participant were years of smoking initiation and smoking cessation (if applicable). From the available data we constructed four variables: 1) For subjects who were smokers in their first wave, we created a binary variable for quitting during the observation period, which we refer to as "recent smoking cessation"; 2) For all individuals who were former smokers at Wave 9, we created a "years since smoking cessation" variable representing years since quitting relative to Wave 9 (<20, 20-<30, ≥30), capturing historical smoking cessation across the lifetime; 3) For ever-smokers, we created a "smoking duration" variable in years (<20, 20-<40, ≥40) using the reported year of smoking initiation and year of smoking cessation (or 2008 for those who still smoked in Wave 9); 4) Data on smoking intensity prior to the study was not collected, but for individuals who smoked during the study, we calculated "on-study" pack-years (<5, 5–<10, ≥10). We imputed missing cigarettes smoked per day for each wave by last observation carried forward. Estimated lifetime pack-years, which was constructed assuming the smoking intensity prior to enrollment was the same as lifetime maximum smoking intensity, was also investigated.

Other covariates were age, sex, race, education, income-to-poverty ratio (i.e., ratio of household income to the poverty threshold for household size), body mass index, alcohol

consumption (0, 1-6, or 7+ drinks per week), and daily physical activity (none, light, medium, vigorous). Covariates were either provided by or derived from RAND HRS Version M (35).

Health conditions related to smoking may also be predictors of TL or consequences of variation in TL. We used two variables as indicators of individuals' overall health status: a co-morbidity index that sums indicators for whether respondents reported a doctor had ever told them they had any of eight conditions: high blood pressure, diabetes, cancer, lung disease, heart disease, stroke, psychiatric problems, and arthritis; and self-reported general health status, with 1 being excellent to 5 being poor.

Statistical analysis

To assess the cross-sectional association between current smoking status and TL in Wave 9, we used a linear regression model with TL as the outcome. Additive interactions between smoking and age/sex were tested. Sex- and age-stratified analyses (<65, 65+) were also performed to assess potential differences in associations by sex and age. Analyses were repeated using prospective measures of smoking status reported at Waves 1-8 rather than smoking status at Wave 9. This analysis was restricted to subjects with TL data who entered in Wave 1 (n=2,369).

Additional smoking characteristics (Wave 9 cigarettes/day, recent smoking cessation, years since cessation, smoking duration, and "on-study" pack-years) were also tested in covariate-adjusted models. To assess the association between TL and overall health, models were fit containing each of the health measures as additional covariates. We also assessed potential residual confounding due to experimental/batch variation in TL measurement by further adjusting the statistical model for DNA sample batch (HRS-provided plate numbers). In

sensitivity analyses we used the post-hoc sampling weights provided by HRS (for each wave based on post-stratification of the sample to the Current Population Survey or American Community Survey for the survey year).

Additional stratification by age, adjustment for survey weights, modeling using normal-quantile transformed TL as the outcome, and adjustment for sample batch were also performed.

All analyses were conducted using Stata 14.0 (StataCorp, College Station, TX).

Simulations to assess bias due to loss to follow-up

To assess potential bias arising from differential loss to follow-up with respect to smoking and short TL, we conducted analyses of simulated datasets (Web Appendix 1).

RESULTS

Cross-sectional analysis

The characteristics of the 5,624 HRS subjects at the time of TL measurement (Wave 9), as well as TL means and standard deviations within strata of participant characteristics are shown in Table 3.1.

5,517 subjects had non-missing data for all demographic and lifestyle covariates included in the cross-sectional regression analysis. Table 3.2 displays the regression coefficients and confidence intervals for all variables included in the multivariate model. Beta coefficients are interpreted as the per unit difference in ln(T/S) for continuous variables or the difference in ln(T/S) compared to reference category for categorical variables, adjusting for all other covariates. TL was shorter with increasing age (P=4.0x10^{-17}), longer among women (P=5.8x10^{-3}), and longer among African Americans and Hispanics compared to white subjects (P=1.7x10^{-11}

Table 3.1. Mean Telomere Length Within Strata of Subject Characteristics, Health and Retirement Study, 2008

Characteristics	Male		Female		All	
	Mean(SD)	n	Mean(SD)	N	Mean(SD)	n
Age, years						
50-54	1.46(0.35)	66	1.42(0.30)	209	1.43(0.31)	275
55-59	1.38(0.36)	355	1.41(0.33)	483	1.39(0.34)	838
60-64	1.34(0.35)	270	1.38(0.33)	462	1.37(0.33)	732
65-69	1.29(0.34)	447	1.34(0.33)	613	1.32(0.34)	1,060
70-74	1.30(0.32)	435	1.30(0.35)	601	1.30(0.34)	1,036
75-79	1.28(0.35)	361	1.29(0.35)	401	1.28(0.35)	762
80-84	1.25(0.33)	223	1.28(0.31)	278	1.27(0.31)	501
85+	1.23(0.39)	153	1.28(0.34)	267	1.26(0.36)	420
Race						
White	1.29(0.34)	1,781	1.32(0.33)	2,430	1.30(0.33)	4,211
African American	1.40(0.41)	266	1.42(0.37)	466	1.42(0.38)	732
Hispanic	1.33(0.34)	213	1.38(0.33)	344	1.36(0.33)	557
Other	1.35(0.38)	50	1.33(0.36)	74	1.34(0.37)	124
Education						
<High School	1.31(0.34)	464	1.31(0.34)	700	1.31(0.34)	1,164
GED	1.23(0.30)	103	1.36(0.36)	142	1.31(0.34)	245
High school graduate	1.28(0.37)	616	1.32(0.33)	1,134	1.31(0.34)	1,750
Some college	1.34(0.37)	501	1.37(0.35)	769	1.36(0.36)	1,270
College+	1.31(0.32)	624	1.36(0.32)	566	1.33(0.32)	1,190
Income-poverty ratio						
<1	1.32(0.32)	129	1.35(0.38)	370	1.34(0.37)	499
1-1.9	1.30(0.36)	359	1.29(0.34)	629	1.29(0.35)	988
2-3.9	1.28(0.34)	735	1.34(0.33)	1,066	1.32(0.34)	1,801
4+	1.32(0.35)	1,087	1.36(0.32)	1,249	1.34(0.34)	2,336
Body mass index (kg/m^2)						
<25	1.26(0.33)	595	1.31(0.33)	1,083	1.29(0.33)	1,678
25-29.9	1.31(0.36)	1,015	1.34(0.34)	1,087	1.33(0.35)	2,102
30-34.9	1.33(0.33)	489	1.35(0.33)	636	1.34(0.33)	1,125
35-39.9	1.35(0.39)	151	1.35(0.32)	292	1.35(0.35)	443
40+	1.36(0.29)	60	1.40(0.38)	216	1.39(0.36)	276
Smoking						
Never smoker	1.33(0.34)	727	1.37(0.33)	1,687	1.36(0.33)	2,414
Former smoker	1.30(0.36)	1,248	1.31(0.35)	1,194	1.30(0.35)	2,442
Current smoker	1.31(0.33)	320	1.29(0.33)	417	1.30(0.33)	737
Alcohol						
0 drinks per week	1.29(0.34)	1,307	1.33(0.34)	2,422	1.32(0.34)	3,729

Table 3.1. Continued

1-7 weeks per week	**1.34(0.36)**	**550**	**1.35(0.34)**	**622**	**1.35(0.35)**	**1,172**
7+ drinks per week	1.32(0.35)	448	1.35(0.33)	266	1.33(0.34)	714
Daily activity						
None	1.30(0.35)	1,913	1.34(0.34)	2,618	1.32(0.34)	4,531
Light	1.37(0.32)	100	1.30(0.34)	306	1.31(0.34)	406
Medium	1.32(0.33)	213	1.32(0.32)	304	1.32(0.33)	517
Vigorous	1.36(0.34)	82	1.41(0.33)	80	1.38(0.34)	162
Total	1.31(0.35)	2,310	1.34(0.34)	3,314	1.32(0.34)	5,624

Abbreviations: GED, General Education Diploma; SD, standard deviation.
Note that some variables have missing values, with a maximum of 31 subjects missing any one variable.

and P=3.5x10^{-3}, respectively). TL was also positively associated with increasing education, income, body mass index, alcohol consumption, and physical activity. Estimates were similar when stratified by age (<65, 65+) (Supplementary Table 3.1), adjusted for survey weights (Supplementary Table 3.2), modeled using normal-quantile transformed TL as the outcome (Supplementary Table 3.3), and adjusted for sample batch (Supplementary Table 3.4).

Among women, we observed shorter TL for both former smokers (P=2.6x10^{-7}) and current smokers (P=6.3x10^{-8}) compared to never smokers (Table 3.2); however, these associations were not observed among men, and a test for interaction indicated that these cross-sectional associations differed by sex ($P_{interaction}$=0.05 and $P_{interaction}$=0.03 for former and current smoking, respectively). Similarly, smoking more cigarettes per day was associated with shorter TL during Wave 9 in women, but not men (Table 3.3).

Analysis of prospective measures of smoking

We examined the associations between Wave 9 TL and smoking status in earlier waves (Figure 3.1 and Supplementary Table 3.5). Associations between smoking measured in early

Table 3.2. Sex-stratified and Overall Associations between Telomere Length and Demographic and Lifestyle Variables, Adjusting for All Variables as Covariates, the Health and Retirement Study, 2008

Characteristics	Male (n = 2,286)		Female (n = 3,231)		All (n = 5,517)	
	β^a	P	β^a	P	β^a	P
Age	-0.003	$5.0x10^{-6}$	-0.003	$3.0x10^{-12}$	-0.003	$4.0x10^{-17}$
Female					0.020	$5.8x10^{-3}$
Race						
White	Ref.		Ref.		Ref.	
African American	0.074	$3.3x10^{-5}$	0.072	$8.3x10^{-8}$	0.072	$1.7x10^{-11}$
Hispanic	0.018	0.37	0.048	$2.6x10^{-3}$	0.037	$3.5x10^{-3}$
Other	0.018	0.63	-0.023	0.44	-0.003	0.91
Education						
<High School	0.005	0.80	-0.025	0.12	-0.011	0.37
GED	-0.060	0.04	0.015	0.55	-0.018	0.33
High-school graduate	-0.030	0.05	-0.009	0.52	-0.016	0.10
Some college	0.005	0.74	0.014	0.31	0.012	0.26
College and above	Ref.		Ref.		Ref.	
Income-poverty ratio						
<1	-0.015	0.56	-0.012	0.46	-0.016	0.27
1-1.9	-0.009	0.63	-0.033	0.01	-0.024	0.02
2-3.9	-0.017	0.21	-0.003	0.76	-0.008	0.34
4+	Ref.		Ref.		Ref.	
Body mass index (kg/m^2)	0.004	$1.5x10^{-4}$	0.001	0.53	0.002	$6.6x10^{-3}$
Smoking						
Never smoker	Ref.		Ref.		Ref.	
Former smoker	-0.023	0.07	-0.049	$2.6x10^{-7}$	-0.039	$2.1x10^{-7}$
Current smoker	-0.019	0.30	-0.078	$6.3x10^{-8}$	-0.056	$7.0x10^{-7}$
Alcohol						
0 drinks per week	Ref.		Ref.		Ref.	
1-7 drinks per week	0.029	0.03	0.012	0.33	0.020	0.02
7+ drinks per week	0.027	0.06	0.029	0.08	0.026	0.02
Daily Activity						
None	Ref.		Ref.		Ref.	
Light	0.052	0.05	-0.046	$2.5x10^{-3}$	-0.018	0.17
Medium	0.024	0.20	-0.008	0.59	0.007	0.58
Vigorous	0.050	0.10	0.033	0.24	0.039	0.05

Abbreviations: GED, General Education Diploma; SD, standard deviation.
[a]Beta coefficients represent the natural log T/S ratio per unit difference for continuous variables, or difference in natural log T/S ratio compared to reference category for categorical variables.

Table 3.3. Sex-stratified and Overall Associations between Telomere Length Measured at Wave 9 and Smoking-Related Characteristics, the Health and Retirement Study, 2008

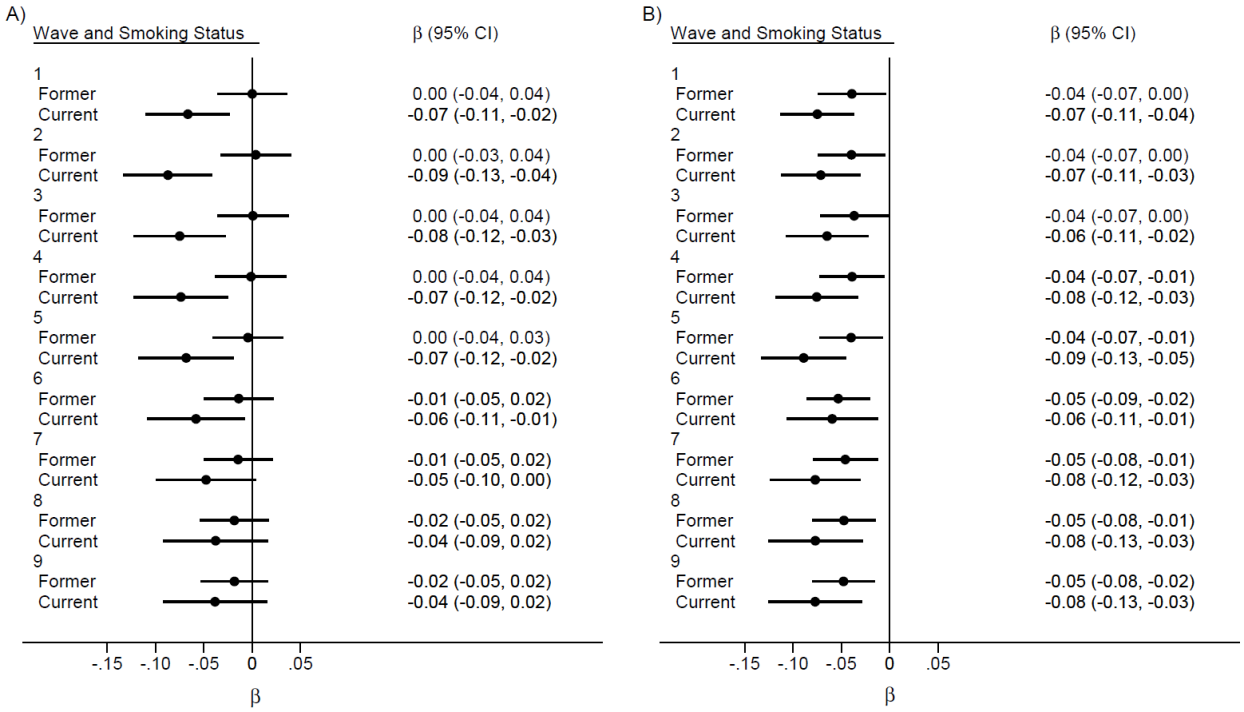
Characteristics	Male				Female				All[a]			
	n	β[b]	P	P_{trend}	n	β[b]	P	P_{trend}	n	β[b]	P	P_{trend}
Wave 9 Smoking quantity (cigarettes/day)												
Never	724	Ref.			1,661	Ref.			2,385	Ref.		
<10	80	-0.057	0.07		127	-0.040	0.08		207	-0.049	$6.4x10^{-3}$	
10-<20	85	0.006	0.84		140	-0.095	$1.2x10^{-5}$		225	-0.056	$1.3x10^{-3}$	
≥20	148	-0.014	0.58	0.64	136	-0.092	$2.9x10^{-3}$	$1.9x10^{-8}$	284	-0.058	$2.9x10^{-4}$	$4.8x10^{-6}$
Recent smoking cessation (after entering study as a smoker)												
Continuing Smoker	320	Ref.			404	Ref.			724	Ref.		
Quitter[c]	152	-0.087	$4.1x10^{-3}$		169	-0.031	0.27		321	-0.048	0.02	
Years since smoking cessation (among former smokers at Wave 9)												
Never	724	Ref.			1,661	Ref.			2,385	Ref.		
≥30	519	0.007	0.64		394	-0.020	0.15		913	-0.007	0.48	
20-<30	284	-0.032	0.08		251	-0.037	0.03		535	-0.037	$2.2x10^{-3}$	
<20	284	-0.055	$2.9x10^{-3}$	$1.3x10^{-3}$	319	-0.076	$4.8x10^{-7}$	$2.0x10^{-7}$	603	-0.067	$8.0x10^{-9}$	$2.5x10^{-9}$
Duration of smoking (years)												
Never	724	Ref.			1,661	Ref.			2,385	Ref.		
<20	413	0.003	0.87		386	-0.010	0.46		799	-0.007	0.53	
20-<40	541	-0.029	0.05		492	-0.060	$2.7x10^{-6}$		1033	-0.049	$3.8x10^{-7}$	
≥40	344	-0.033	0.06	0.018	364	-0.088	$1.8x10^{-9}$	$6.2x10^{-12}$	708	-0.064	$1.1x10^{-8}$	$1.9x10^{-11}$
Pack-years during the study (sum of packs/day*years smoked per wave)[d]												
Never	724	Ref.			1,661	Ref.			2,385	Ref.		
<5	233	-0.039	0.06		323	-0.081	$1.6x10^{-7}$		556	-0.064	$1.9x10^{-7}$	

Table 3.3. Continued

5-<10	124	-0.058	0.02		168	-0.080	$1.1x10^{-4}$		292	-0.070	$1.2x10^{-5}$	
≥10	182	-0.060	$8.6x10^{-3}$	$2.8x10^{-3}$	166	-0.080	$1.1x10^{-4}$	$7.5x10^{-10}$	348	-0.074	$1.2x10^{-6}$	$3.4x10^{-11}$

[a]Missing data among subjects with telomere length measurement was <1% (31/5,554) for Wave 9 smoking status, 1% (8/724) for Wave 9 smoking rate, 15% (366/2,414) for years since smoking cessation, 18% (565/3138) for smoking duration, and 2% (28/1243) for "on-study" pack-years. Telomere length did not differ by missing status of years since smoking cessation (P=0.41), or smoking duration (P= 0.69).
[b]Estimates are from a linear regression of natural log of T/S ratio on smoking characteristics adjusted for sex (when appropriate), age, race, education, income, body mass index, alcohol consumption, and daily physical activity.
[c]Subjects who quit smoking between wave of entry and Wave 9.
[d]Estimates are further adjusted for wave of entry into study.

Figure 3.1. Beta coefficients and 95% confidence intervals for the associations between telomere length measured at Wave 9 in the Health and Retirement Study (2008) and smoking status measured across multiple waves among men (A) and women (B), adjusting for age, race, education, income, body mass index, alcohol consumption, and daily physical activity. Samples were restricted to subjects who entered the study in Wave 1 and were measured for telomere length at Wave 9 (male n = 1,092, female n = 1,277). Beta coefficients represent the difference in natural log T/S ratio compared to never smokers.

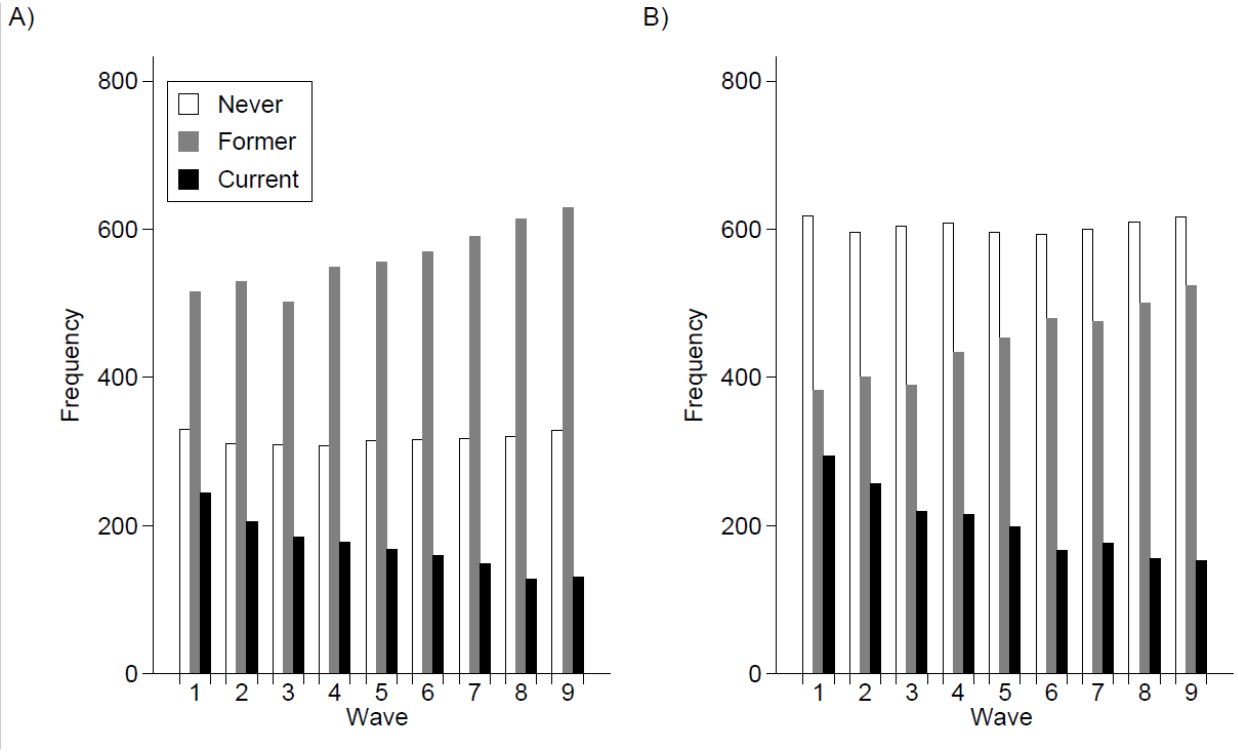


waves and subsequent TL were similar for men (Figure 3.1A) and women (Figure 3.1B). Among men, we observed a gradual attenuation of the association in later waves. The point estimate decreased in magnitude by more than 50%, with *P*-values ranging from $P=2x10^{-4}$ to $P=0.17$, consistent with a weaker association for smoking reported in more recent waves. For women, the association between shorter TL and both current and former smoking status was consistent across all waves.

Additional characteristics of smoking

In order to determine why the association between smoking and TL among men was weaker for more recent measures of smoking, we first examined the change in distribution of smoking status over time. We observed that almost half of those in the current smoker category in Wave 1 moved to the former smoking category by Wave 9 among both men (Figure 3.2A) and women (Figure 3.2B). However, despite the similar rates of quitting, we observed a key difference between men and women who quit: Male smokers who quit during the study had shorter TL compared to those who continued to smoke ($P=4.1x10^{-3}$) (Table 3.3); this differential quitting by TL was not observed among female smokers ($P=0.27$). Conversely, when assessing other characteristics of smoking behavior such as years since cessation (among former smokers), duration, and "on-study" pack-years, we observed consistent evidence that greater smoking exposure was associated with shorter TL in both men and women (Table 3.3). Analyses of estimates of life-time pack-years showed similar results (Supplementary Table 3.6).

Figure 3.2. Distribution of smoking statuses among men (A) and women (B) across Waves 1-9, from the Health and Retirement Study (1992-2008). Samples were restricted to subjects who entered the study in Wave 1 and were measured for telomere length at Wave 9 (men n = 1,095, women n = 1,301).



Analysis of TL and overall health status

In order to further understand the link between smoking cessation and short TL in men, we examined the association between TL and health status. In the analyses of TL and measures of health status, we observed shorter TL among men who reported poorer health based on self-reported general health status (P=0.02) and the co-morbidity index (P=0.05), but these associations were not observed among women (Table 3.4).

Simulations to assess bias due to loss to follow-up

Simulations demonstrated that when there was an increased risk of loss to follow-up both for smokers and individuals with shorter TL, the association between smoking and TL was biased in a positive direction (Supplementary Table 3.7). Therefore, under the assumption that both smoking and short TL increase the probability of being lost to follow-up prior to Wave 9 (presumably through increased morbidity and mortality), the associations between smoking and short TL observed in this study would be biased towards the null.

DISCUSSION

In this study of older men and women in the U.S., we observed that smoking was associated with TL when using prospective measures of smoking status taken up to 16 years prior to TL measurement, supporting the hypothesis that smoking reduces TL. However, this association was strongly attenuated in cross-sectional analyses among men but not women. Our analyses indicated that the attenuation was driven by higher rates of male smokers with shorter TL quitting smoking during the course of the study compared to male smokers with longer TL. The movement of men with shorter TL out of "current smoking" and into "former smoking" led

Table 3.4. Associations between Telomere Length and Health Conditions, Stratified by Sex, From the Health and Retirement Study, 2008

Health Status Measure	Male (n=2,284)		Female (n=3,232)		All (n=5,516)	
	β[a]	P	β[a]	P	β[a]	P
Self-reported health[b]	-0.009	0.02	0.002	0.53	-0.003	0.28
Sum of conditions[c]	-0.011	0.05	0.003	0.50	-0.002	0.48

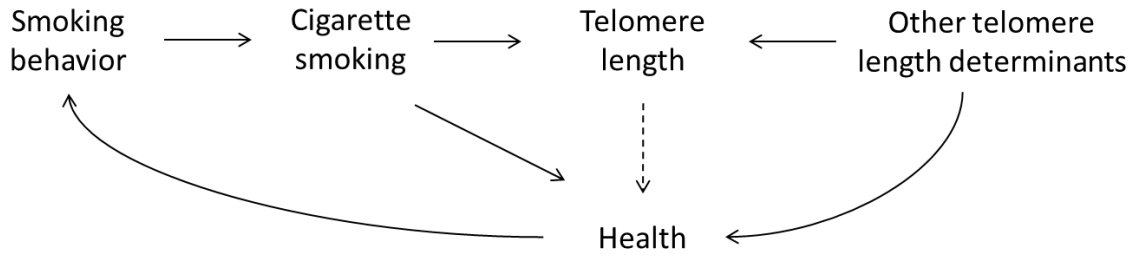
[a]Estimates are from a linear regression adjusted for sex (when appropriate), age, race, education, income, body mass index, smoking status, alcohol consumption, and daily physical activity.
[b]Self-reported general health status, with 1 being excellent and 5 being poor
[c]Sum of indicators for whether respondent reported a doctor had ever told them they had any of eight conditions: high blood pressure, diabetes, cancer, lung disease, heart disease, stroke, psychiatric problems, and arthritis.

to an overrepresentation of men with longer TL in the "current smoking" category. This phenomenon resulted in very weak associations between TL and smoking status in later waves, with those associations likely biased estimates of the true effect of smoking on TL.

Male smokers with shorter TL may be more likely to quit because short TL was correlated with poorer health for men in HRS; furthermore, previous studies from HRS show that smokers who experience acute and chronic health events are more likely to quit smoking than individuals who do not (36,37). This may be an example of the "healthy smoker" phenomenon in which individuals who are more resistant to the adverse effects of smoking continue to smoke while those who are more susceptible quit, resulting in underestimation of the effect of smoking (38). This adverse health-driven smoking cessation may be particularly relevant to men in HRS because they are more likely to experience major and chronic health events (36), and to older individuals, who are more likely than younger individuals to experience adverse health events and to quit smoking in response (39).

Our hypothesized relationships among smoking, TL, and health for men are shown in Figure 3.3. The bias towards the null observed in cross-sectional analyses of men can be conceptualized in several ways. First, assuming TL affects health, this bias could be attributed to

Figure 3.3. Causal model showing the hypothesized relationships among smoking, telomere length, health outcomes, and unmeasured determinants of telomere length in male HRS participants. Solid lines depict associations that are likely to represent causal relationships. The dashed line depicts a potential effect.



reverse causation, whereby TL influences smoking behavior through its effects on health (Figure 3.3). Second, we could view this bias as confounding of the smoking-TL association by factors that influence both TL and health (and health-related smoking behaviors) in later life (Figure 3.3). Third, the bias could be considered self-selection bias, in which individuals select themselves into a group (i.e., former smokers) in a non-random way (i.e., based on TL and health). Fourth, we could consider current smoking to be an error-prone measure of the true exposure of interest (e.g., lifetime exposure or exposure in early- to mid-life), thus producing bias due to differential measurement error (differential with respect to TL).

In contrast to men, the association between smoking status and TL did not vary across waves among women, and female smokers who quit during the study did not exhibit shorter TL than smokers who did not quit. This could be due to different motivations between men and women for quitting smoking (40), i.e. women are less likely to quit in response to an adverse health event than men (41). Also, women tend to be less successful than men with regard to smoking cessation (42,43), and therefore may be less able to maintain cessation in response to an adverse health event. Finally, sex differences in the accuracy of self-reported smoking behavior, exposure to second-hand smoking, duration and intensity of smoking, and biological effects of smoking on TL could all contribute to the differences we observed.

We utilized two different sources of data to study smoking cessation in our analysis – the act of quitting during the study period and the reported time since quitting during the individual's lifetime. The former represents recent quitting among individuals smoking in later life (which may correlate with concurrent disease status and health-seeking behavior) while the latter captures smoking cessation events relevant to cumulative lifetime exposure. The association observed between greater time since smoking cessation and longer TL was consistent with the

results from the prospective analysis, as well as analyses of two other characteristics of smoking exposure, i.e., duration and pack-years, supporting the hypothesis that smoking shortens telomeres.

This is the first study to characterize the association between smoking and subsequent TL while treating smoking as a time-varying exposure. The differences in this association that we observed across time and between sexes potentially explain conflicting findings in prior cross-sectional, longitudinal, and prospective studies of smoking and TL. Previous cross-sectional studies reporting an association between smoking and short TL tend to be studies of female subjects only (11–14,16), while those reporting no association tend to be studies of men or of older subjects (21,23–26), where estimates may be more likely to be biased due to smoking cessation among older men with short TL. Previous longitudinal studies reporting associations between smoking and TL attrition involved younger subjects compared to those not reporting such an association (17–20), further suggesting that estimates obtained from studies of older individuals could be biased due to adverse health-driven smoking cessation. The analysis of the Helsinki Businessmen Study compared prospective and cross-sectional measures of smoking status and TL among men and reported findings consistent with our study (27). However, that study assessed smoking at only two time points, and did not consider smoking cessation as an explanation for the phenomenon, instead positing that unmeasured factors contributing to TL attrition in old age such as frailty, telomerase insufficiency, and oxidative stress were possible explanations.

Our study has a number of strengths. A large sample size and rich covariate data allowed for comprehensive adjustments for demographic and lifestyle variables. The study was conducted within a nationally representative cohort of older subjects—an important age group to

study given the tremendous interest in TL as a biomarker of aging-related diseases and mortality. The use of repeated measures of smoking over 16 years shed light on the dynamics of the association between smoking and TL over time, and demonstrated the importance of evaluating both recent and historical exposure to smoking in order to understand its association with TL over the life course.

Our study is limited by the possible issue of bias due to loss to follow-up. In other words, subjects who were not lost to follow-up (including death) prior to Wave 9 (allowing for TL measurement) may be systematically different from those who were lost to follow-up by Wave 9 with respect to the variables analyzed, and this selection could bias the association between smoking and TL. This phenomenon has been previously described as a type of collider bias resulting from selection into a study conditional on a variable (e.g. survival, health) that is influenced by other variables under analysis (e.g. smoking and TL) (44). We conducted simulation analyses demonstrating that greater loss to follow-up among male smokers and among men with short TL would create a positive bias that would increase the beta coefficient for the association between smoking and TL. Such a bias would therefore attenuate the association between smoking and short TL. Such a bias could contribute, in part, to the lack of association between smoking and TL observed in cross-sectional analyses among men. However, our analyses of individuals present in both Wave 1 and Wave 9 (1,095 men not lost to follow-up) showed significant associations between prior smoking (at Wave 1) and short TL (at Wave 9) despite the presence of any such attenuating bias. Cross-sectional analyses of these same 1,095 male participants showed no smoking-TL association, but that attenuation was apparently due to a higher rate of quitting among individuals with short TL. Taken together, these findings indicated that 1) the association between prospectively-measured smoking and subsequent short

TL was observed *despite* potential bias due to loss to follow-up and 2) the cross-sectional null association was not primarily driven by bias due to loss to follow-up.

Another limitation is that it was necessary to assume that the available TL measure at Wave 9 was relatively constant within the time scale of this study and therefore an appropriate proxy for TL across earlier waves when quitting behavior occurred. This assumption is likely valid as prior findings indicate that there is a weak association between recent (~12 years) changes in smoking behavior and TL attrition (45). Finally, we note that the source of DNA in this study, saliva, is different from most previous studies of TL which use leukocyte DNA. Although up to 74% of saliva DNA originates from white blood cells, the composition of leukocytes in normal saliva samples will differ from the composition of blood samples (46). However, previous studies have shown that the quality of saliva genomic DNA are comparable with blood samples based on DNA purity, genotyping, and polymerase chain reaction amplification analyses (47,48), and TL measured from saliva DNA has been found to be highly correlated with TL measured from blood leukocyte DNA (r=0.72) (49). Additionally, there is no *a priori* reason to favor leukocyte over saliva DNA source to assess the association between smoking and TL in human cells in the absence of a specific target tissue of interest for health outcome assessment.

In summary, we used longitudinal data on a representative sample of older adults in the U.S. to demonstrate that smoking is associated with reduced TL among both men and women when using repeated measures of smoking status over time. However, in cross-sectional analyses, smoking status was not associated with TL among men, and this apparent discrepancy was likely driven by a higher propensity for smoking cessation among men with short TL, a condition which appears to be accompanied by poorer overall health. These findings provide a framework

for understanding conflicting reports of smoking and TL in previous studies and highlight the need to further investigate the time-varying and sex-specific associations among smoking behavior, TL and health.

ACKNOWLEDGMENTS

This work was supported by the National Institute on Aging (T32AG000243; P30AG012857), the National Institute of Environmental Health (R01 ES020506), and the National Human Genome Research Institute (U01 HG007601) at the National Institutes of Health.

REFERENCES

1. Müezzinler A, Zaineddin AK, Brenner H. A systematic review of leukocyte telomere length and age in adults. Ageing Res Rev. 2013;12:509–19.

2. Sanders JL, Newman AB. Telomere Length in Epidemiology: A Biomarker of Aging, Age-Related Disease, Both, or Neither? Epidemiol Rev. 2013;35:112–31.

3. Serrano AL, Andrés V. Telomeres and Cardiovascular Disease: Does Size Matter? Circ Res. 2004;94:575–84.

4. Willeit P, Willeit J, Mayr A, Weger S, Oberhollenzer F, Brandstätter A, et al. Telomere length and risk of incident cancer and cancer mortality. J Am Med Assoc. 2010;304:69–75.

5. Martin-Ruiz C, Dickinson HO, Keys B, Rowan E, Kenny RA, Von Zglinicki T. Telomere length predicts poststroke mortality, dementia, and cognitive decline. Ann Neurol. 2006;60:174–80.

6. van der Vaart H, Postma DS, Timens W, Ten Hacken NHT. Acute effects of cigarette smoke on inflammation and oxidative stress: a review. Thorax. 2004;59:713–21.

7. Von Zglinicki T. Oxidative stress shortens telomeres. Trends Biochem Sci. 2002;27:339–344.

8. O'Donovan A, Pantell MS, Puterman E, Dhabhar FS, Blackburn EH, Yaffe K, et al. Cumulative Inflammatory Load Is Associated with Short Leukocyte Telomere Length in the Health, Aging and Body Composition Study. PLoS ONE. 2011;6:e19687.

9. Jenny NS. Inflammation in aging: cause, effect, or both? Discov Med. 2012;13:451–60.

10. Babizhayev MA, Yegorov YE. Smoking and health: association between telomere length and factors impacting on human disease, quality of life and life span in a large population-based cohort under the effect of smoking duration. Fundam Clin Pharmacol. 2011;25:425–42.

11. Carty CL, Kooperberg C, Liu J, Herndon M, Assimes T, Hou L, et al. Leukocyte Telomere Length and Risks of Incident Coronary Heart Disease and Mortality in a Racially Diverse Population of Postmenopausal Women. Arterioscler Thromb Vasc Biol. 2015;35:2225–31.

12. Aviv A, Valdes A, Gardner JP, Swaminathan R, Kimura M, Spector TD. Menopause Modifies the Association of Leukocyte Telomere Length with Insulin Resistance and Inflammation. J Clin Endocrinol Metab. 2006;91:635–40.

13. Valdes A, Andrew T, Gardner J, Kimura M, Oelsner E, Cherkas L, et al. Obesity, cigarette smoking, and telomere length in women. The Lancet. 20;366:662–4.

14. Parks CG, Miller DB, McCanlies EC, Cawthon RM, Andrew ME, DeRoo LA, et al. Telomere length, current perceived stress, and urinary stress hormones in women. Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol. 2009;18:551–60.

15. O'Donnell CJ, Demissie S, Kimura M, Levy D, Gardner JP, White C, et al. Leukocyte Telomere Length and Carotid Artery Intimal Medial Thickness: The Framingham Heart Study. Arterioscler Thromb Vasc Biol. 2008;28:1165–71.

16. Sun Q, Shi L, Prescott J, Chiuve SE, Hu FB, De Vivo I, et al. Healthy Lifestyle and Leukocyte Telomere Length in U.S. Women. PLoS ONE. 2012;7:e38374.

17. Huzen J, Wong LSM, van Veldhuisen DJ, Samani NJ, Zwinderman AH, Codd V, et al. Telomere length loss due to smoking and metabolic traits. J Intern Med. 2014;275:155–63.

18. Bendix L, Thinggaard M, Fenger M, Kolvraa S, Avlund K, Linneberg A, et al. Longitudinal Changes in Leukocyte Telomere Length and Mortality in Humans. J Gerontol A Biol Sci Med Sci. 2014;69A:231–9.

19. Weischer M, Bojesen SE, Nordestgaard BG. Telomere shortening unrelated to smoking, body weight, physical activity, and alcohol intake: 4,576 general population individuals with repeat measurements 10 years apart. PLoS Genet. 2014;10.

20. Müezzinler A, Mons U, Dieffenbach AK, Butterbach K, Saum K-U, Schick M, et al. Smoking habits and leukocyte telomere length dynamics among older adults: Results from the ESTHER cohort. Exp Gerontol. 2015;70:18–25.

21. Bischoff C, Graakjaer J, Petersen HC, Hjelmborg J v B, Vaupel JW, Bohr V, et al. The heritability of telomere length among the elderly and oldest-old. Twin Res Hum Genet Off J Int Soc Twin Stud. 2005;8:433–9.

22. Bekaert S, De Meyer T, Rietzschel ER, De Buyzere ML, De Bacquer D, Langlois M, et al. Telomere length and cardiovascular risk factors in a middle-aged population free of overt cardiovascular disease. Aging Cell. 2007;6:639–47.

23. Fitzpatrick AL, Kronmal RA, Gardner JP, Psaty BM, Jenny NS, Tracy RP, et al. Leukocyte Telomere Length and Cardiovascular Disease in the Cardiovascular Health Study. Am J Epidemiol. 2007;165:14–21.

24. Nordfjäll K, Eliasson M, Stegmayr B, Melander O, Nilsson P, Roos G. Telomere Length Is Associated With Obesity Parameters but With a Gender Difference. Obesity. 2008;16:2682–9.

25. Batty GD, Wang Y, Brouilette SW, Shiels P, Packard C, Moore J, et al. Socioeconomic status and telomere length: the West of Scotland Coronary Prevention Study. J Epidemiol Community Health. 2009;63:839–41.

26. Brouilette SW, Moore JS, McMahon AD, Thompson JR, Ford I, Shepherd J, et al. Telomere length, risk of coronary heart disease, and statin treatment in the West of Scotland Primary Prevention Study: a nested case-control study. The Lancet. 13;369:107–14.

27. Strandberg TE, Saijonmaa O, Tilvis RS, Pitkälä KH, Strandberg AY, Miettinen TA, et al. Association of telomere length in older men with mortality and midlife body mass index and smoking. J Gerontol A Biol Sci Med Sci. 2011;66:815–20.

28. Barrett ELB, Richardson DS. Sex differences in telomeres and lifespan. Aging Cell. 2011;10:913–21.

29. Aviv A, Shay J, Christensen K, Wright W. The longevity gender gap: are telomeres the explanation? Sci Aging Knowl Environ SAGE KE. 2005;2005:pe16.

30. Aviv A. Telomeres, sex, reactive oxygen species, and human cardiovascular aging. J Mol Med. 2002;80:689–95.

31. Bolego C, Poli A, Paoletti R. Smoking and gender. Cardiovasc Res. 2002;53:568–76.

32. Bauer T, Göhlmann S, Sinning M. Gender differences in smoking behavior. Health Econ. 2007;16:895–909.

33. Sonnega A, Faul JD, Ofstedal MB, Langa KM, Phillips JW, Weir DR. Cohort Profile: the Health and Retirement Study (HRS). Int J Epidemiol. 2014;43:576–85.

34. Health and Retirement Study, HRS 1992-2008 public use datasets, 2008 telomere length data. Produced and distributed by the University of Michigan with funding from the National Institute on Aging (grant number NIA U01AG009740). Ann Arbor, MI. University of Michigan Web site. http://hrsonline.isr.umich.edu/modules/meta/telo2008/desc/Telomere08DD.pdf. Published December 2013. Accessed February 16, 2016.

35. Chien S, Campbell N, Hayden O, Hurd M, Main R, Mallett J, et al. RAND HRS data documentation, version M. Technical report, Labor & Population Program, RAND Center for the Study of Aging; 2013.

36. Falba T. Health Events and the Smoking Cessation of Middle Aged Americans. J Behav Med. 2005;28:21–33.

37. Keenan PS. Smoking and weight change after new health diagnoses in older adults. Arch Intern Med. 2009;169:237–42.

38. Becklake MR, Lalloo U. The "healthy smoker": a phenomenon of health selection? Respir Int Rev Thorac Dis. 1990;57:137–44.

39. Grøtvedt L, Stavem K. Association between age, gender and reasons for smoking cessation. Scand J Public Health. 2005;33:72–6.

40. McKee SA, O'Malley SS, Salovey P, Krishnan-Sarin S, Mazure CM. Perceived risks and benefits of smoking cessation: Gender-specific predictors of motivation and treatment outcome. Addict Behav. 2005;30:423–35.

41. McKee SA, Maciejewski PK, Falba T, Mazure CM. Sex differences in the effects of stressful life events on changes in smoking status. Addict Abingdon Engl. 2003;98:847–55.

42. Wetter DW, Kenford SL, Smith SS, Fiore MC, Jorenby DE, Baker TB. Gender differences in smoking cessation. J Consult Clin Psychol. 1999;67:555–62.

43. Piper ME, Cook JW, Schlam TR, Jorenby DE, Smith SS, Bolt DM, et al. Gender, race, and education differences in abstinence rates among participants in two randomized smoking cessation trials. Nicotine Tob Res Off J Soc Res Nicotine Tob. 2010;12:647–57.

44. Pearce N, Richiardi L. Commentary: Three worlds collide: Berkson's bias, selection bias and collider bias. Int J Epidemiol. 2014;43:521–4.

45. Benetos A, Kark JD, Susser E, Kimura M, Sinnreich R, Chen W, et al. Tracking and fixed ranking of leukocyte telomere length across the adult life course. Aging Cell. 2013;12:615–21.

46. Thiede C, Prange-Krex G, Freiberg-Richter J, Bornhäuser M, Ehninger G. Buccal swabs but not mouthwash samples can be used to obtain pretransplant DNA fingerprints from recipients of allogeneic bone marrow transplants. Bone Marrow Transplant. 2000;25:575–7.

47. Rylander-Rudqvist T, Håkansson N, Tybring G, Wolk A. Quality and Quantity of Saliva DNA Obtained from the Self-administrated Oragene Method—A Pilot Study on the Cohort of Swedish Men. Cancer Epidemiol Biomarkers Prev. 2006;15:1742–5.

48. Hansen T v. O, Simonsen MK, Nielsen FC, Hundrup YA. Collection of Blood, Saliva, and Buccal Cell Samples in a Pilot Study on the Danish Nurse Cohort: Comparison of the Response Rate and Quality of Genomic DNA. Cancer Epidemiol Biomarkers Prev. 2007;16:2072–6.

49. Mitchell C, Hobcraft J, McLanahan SS, Siegel SR, Berg A, Brooks-Gunn J, et al. Social disadvantage, genetic sensitivity, and children's telomere length. Proc Natl Acad Sci. 2014;111:5944–9.

SUPPLEMENTARY MATERIALS

Simulations to assess bias due to loss to follow-up

To assess the impact of potential bias arising from differential loss to follow-up with respect to smoking and short TL, we conducted analyses of simulated datasets. For each simulated scenario, we generated 1,000 datasets consisting of 5,000 observations and three variables: current smoking status (X), TL (Y), and a binary variable representing loss to follow-up (Z) whose value was influenced by both X and Y. X was randomly generated as a binary variable with a 0.25 probability of being a smoker (x=1) and 0.75 probability of being a non-smoker (x=0) using a Bernoulli distribution. Y was modeled as a random number from a standard normal distribution plus a linear effect of X:

$$y_i = \beta_{yx} x_i + \rho_i \quad with \quad \rho_i \sim N(0,1). \qquad (1)$$

β_{yx} was set to either 0 (no effect of smoking on TL) or -0.2 (smoking reduces TL). The binary variable representing loss to follow-up (Z) was generated as a binary variable using a logistic model:

$$P(Z = 1|X_i, Y_i) = 1/(1 + e^{-(\beta_0 + \beta_{zx} x_i + \beta_{zy} y_i)}) \qquad (2)$$

We simulated β_0 to produce an average population risk of loss to follow-up among non-smokers to be 35% (β_0 = -ln((1-0.35)/0.35)), similar to what is observed in HRS. The effect of smoking on loss to follow up (β_{zx}) was set to 1, 1.5, and 2, and the effect of TL on loss to follow-up (β_{zy}) was set to 1, 0.75, and 0.5, representing an increasing risk of loss to follow-up for smokers and individuals with shorter TL. In each simulated dataset, the association between X and Y was estimated using linear regression for 1) the entire simulated sample and 2) restricting to the individuals who were not lost to follow-up.

Supplementary Table 3.1. Associations Between Telomere Length and Demographic and Lifestyle Variables, Stratified by Age (<65, 65+), Health and Retirement Study, 2008

		<65 (n = 1,807)		65+ (n = 3,711)	
		β	P	β	P
Age (by decade)		-0.037	0.020	-0.02	0.003
Female		0.025	0.036	0.018	0.050
Race	White	Ref.		Ref.	
	African American	0.074	$7.0x10^{-6}$	0.073	$1.2x10^{-7}$
	Hispanic	0.059	0.002	0.017	0.307
	Other	0.061	0.053	-0.064	0.056
Education	<High School	0.007	0.754	-0.014	0.359
	GED	-0.016	0.58	-0.011	0.63
	High-school graduate	-0.026	0.109	-0.01	0.418
	Some college	0.004	0.809	0.017	0.224
	College and above	Ref.		Ref.	
Income-poverty ratio	<1	-0.004	0.841	-0.023	0.201
	1-1.9	-0.016	0.396	-0.027	0.032
	2-3.9	-0.021	0.128	-0.004	0.727
	4+	Ref.		Ref.	
BMI (per 10 units)	BMI	0.011	0.236	0.023	0.005
Smoking	Never smoker	Ref.		Ref.	
	Former smoker	-0.039	0.002	-0.039	$3.7x10^{-5}$
	Current smoker	-0.071	$1.1x10^{-5}$	-0.045	0.004
Alcohol	0 drinks per week	Ref.		Ref.	
	1-7 drinks per week	0.010	0.441	0.023	0.047
	7+ drinks per week	0.019	0.26	0.029	0.034
Daily activity	None	Ref.		Ref.	
	Light	0.002	0.935	-0.031	0.070
	Medium	0.034	0.096	-0.005	0.728
	Vigorous	0.036	0.225	0.040	0.139

Abbreviations: BMI, body mass index.

Supplementary Table 3.2. Survey-Weight Adjusted[a] Associations Between Telomere Length and Demographic and Lifestyle Variables, Stratified by Sex and Pooled, Health and Retirement Study, 2008

		Male (n = 2,286)		Female (n = 3,231)		All (n = 5,517)	
		β	P	β	P	β	P
Age (by decade)		-0.121	$7.5x10^{-7}$	-0.145	$1.4x10^{-14}$	-0.137	$3.7x10^{-20}$
Female						0.084	0.003
Race	White	Ref.		Ref.		Ref.	
	African American	0.308	$7.6x10^{-6}$	0.315	$1.3x10^{-9}$	0.31	$7.6x10^{-14}$
	Hispanic	0.106	0.176	0.201	0.001	0.166	0.001
	Other	0.108	0.451	-0.054	0.644	0.027	0.765
Education	<High School	0.006	0.930	-0.079	0.208	-0.035	0.453
	GED	-0.2	0.069	0.06	0.520	-0.053	0.452
	High-school graduate	-0.098	0.102	-0.032	0.543	-0.054	0.163
	Some college	0.033	0.585	0.071	0.195	0.061	0.133
	College and above	Ref.		Ref.		Ref.	
Income-poverty ratio	<1	-0.046	0.643	-0.046	0.477	-0.054	0.319
	1-1.9	-0.013	0.848	-0.13	0.012	-0.086	0.034
	2-3.9	-0.056	0.275	-0.012	0.775	-0.027	0.401
	4+	Ref.		Ref.		Ref.	
BMI (per 10 units)	BMI	0.162	$1.8x10^{-4}$	0.013	0.635	0.061	0.010
Smoking	Never smoker	Ref.		Ref.		Ref.	
	Former smoker	-0.085	0.076	-0.202	$6.1x10^{-8}$	-0.156	$9.3x10^{-8}$
	Current smoker	-0.069	0.327	-0.315	$1.4x10^{-8}$	-0.224	$2.7x10^{-7}$
Alcohol	0 drinks per week	Ref.		Ref.		Ref.	
	1-7 drinks per week	0.117	0.024	0.032	0.480	0.073	0.033
	7+ drinks per week	0.091	0.105	0.123	0.059	0.096	0.022
Daily activity	None	Ref.		Ref.		Ref.	
	Light	0.201	0.051	-0.142	0.015	-0.046	0.371
	Medium	0.085	0.239	-0.029	0.621	0.024	0.604
	Vigorous	0.173	0.125	0.124	0.259	0.142	0.071

Abbreviations: BMI, body mass index.

[a]Since telomere length was only measured in a subset of the HRS participants, it is unclear how appropriate the HRS post-hoc sampling weights are for this analysis. However, subjects with telomere length appear to be representative of the larger cohort with respect to age and race, suggesting that the provided weights may be appropriate for the analysis.

Supplementary Table 3.3. Associations Between Normal-Quantile Transformed Telomere Length and Demographic and Lifestyle Variables, Stratified by Sex and Pooled, Health and Retirement Study, 2008

		Male (n = 2,286)		Female (n = 3,231)		All (n = 5,517)	
		β	P	β	P	β	P
Age (by decade)		-0.025	0.001	-0.035	$4.2x10^{-8}$	-0.031	$1.4x10^{-8}$
Female						0.028	0.002
Race	White	Ref.		Ref.		Ref.	
	African American	0.055	0.035	0.078	$2.9x10^{-5}$	0.068	$9.2x10^{-6}$
	Hispanic	0.041	0.076	0.029	0.263	0.033	0.096
	Other	0.068	0.176	-0.027	0.491	0.031	0.264
Education	<High School	0.013	0.609	-0.01	0.582	0.003	0.845
	GED	-0.099	0.036	0.039	0.142	-0.026	0.408
	High-school graduate	-0.043	0.063	-0.019	0.269	-0.029	0.031
	Some college	-0.003	0.887	0.018	0.251	0.01	0.402
	College and above	Ref.		Ref.		Ref.	
Income-poverty ratio	<1	-0.052	0.161	-0.015	0.493	-0.028	0.208
	1-1.9	0.010	0.636	-0.028	0.073	-0.013	0.256
	2-3.9	-0.021	0.257	-0.012	0.416	-0.013	0.229
	4+	Ref.		Ref.		Ref.	
BMI (per 10 units)	BMI	0.046	$1.6x10^{-5}$	$3.0x10^{-4}$	0.973	0.018	0.006
Smoking	Never smoker	Ref.		Ref.		Ref.	
	Former smoker	-0.020	0.187	-0.05	$5.9x10^{-6}$	-0.037	$8.9x10^{-6}$
	Current smoker	-0.006	0.839	-0.082	$1.2x10^{-4}$	-0.05	0.004
Alcohol	0 drinks per week	Ref.		Ref.		Ref.	
	1-7 drinks per week	0.036	0.032	0.009	0.554	0.024	0.027
	7+ drinks per week	0.031	0.020	0.011	0.590	0.019	0.128
Daily activity	None	Ref.		Ref.		Ref.	
	Light	0.041	0.070	-0.036	0.030	-0.01	0.482
	Medium	0.057	0.003	-0.002	0.887	0.026	0.036
	Vigorous	0.045	0.053	0.027	0.266	0.034	0.033

Abbreviations: BMI, body mass index.

Supplementary Table 3.4. DNA Sample Batch[a]-Adjusted Associations Between Telomere Length and Demographic and Lifestyle Variables, Stratified by Sex and Pooled, Health and Retirement Study, 2008

		Male (n = 2,286)		Female (n = 3,231)		All (n = 5,517)	
		β	P	β	P	β	P
Age (by decade)		-0.025	0.001	-0.035	$4.2x10^{-8}$	-0.031	$1.4x10^{-8}$
Female						0.028	0.002
Race	White	Ref.		Ref.		Ref.	
	African American	0.055	0.035	0.078	$2.9x10^{-5}$	0.068	$9.2x10^{-6}$
	Hispanic	0.041	0.076	0.029	0.263	0.033	0.096
	Other	0.068	0.176	-0.027	0.491	0.031	0.264
Education	<High School	0.013	0.609	-0.01	0.582	0.003	0.845
	GED	-0.099	0.036	0.039	0.142	-0.026	0.408
	High-school graduate	-0.043	0.063	-0.019	0.269	-0.029	0.031
	Some college	-0.003	0.887	0.018	0.251	0.01	0.402
	College and above	Ref.		Ref.		Ref.	
Income-poverty ratio	<1	-0.052	0.161	-0.015	0.493	-0.028	0.208
	1-1.9	0.010	0.636	-0.028	0.073	-0.013	0.256
	2-3.9	-0.021	0.257	-0.012	0.416	-0.013	0.229
	4+	Ref.		Ref.		Ref.	
BMI (per 10 units)	BMI	0.046	$1.6x10^{-5}$	$3.0x10^{-4}$	0.973	0.018	0.006
Smoking	Never smoker	Ref.		Ref.		Ref.	
	Former smoker	-0.020	0.187	-0.05	$5.9x10^{-6}$	-0.037	$8.9x10^{-6}$
	Current smoker	-0.006	0.839	-0.082	$1.2x10^{-4}$	-0.05	0.004
Alcohol	0 drinks per week	Ref.		Ref.		Ref.	
	1-7 drinks per week	0.036	0.032	0.009	0.554	0.024	0.027
	7+ drinks per week	0.031	0.020	0.011	0.590	0.019	0.128
Daily activity	None	Ref.		Ref.		Ref.	
	Light	0.041	0.070	-0.036	0.030	-0.01	0.482
	Medium	0.057	0.003	-0.002	0.887	0.026	0.036
	Vigorous	0.045	0.053	0.027	0.266	0.034	0.033

Abbreviations: BMI, body mass index.
[a]HRS-provided plate numbers

Supplementary Table 3.5. Associations Between Telomere Length and Smoking Status, Stratified by Sex, Health and Retirement Study, 2008

		Male (n = 2,286)		Female (n = 3,231)		All (n = 5,517)	
		β	P	β	P	β	P
Age (by decade)		-0.025	0.001	-0.035	$4.2x10^{-8}$	-0.031	$1.4x10^{-8}$
Female						0.028	0.002
Race	White	Ref.		Ref.		Ref.	
	African American	0.055	0.035	0.078	$2.9x10^{-5}$	0.068	$9.2x10^{-6}$
	Hispanic	0.041	0.076	0.029	0.263	0.033	0.096
	Other	0.068	0.176	-0.027	0.491	0.031	0.264
Education	<High School	0.013	0.609	-0.01	0.582	0.003	0.845
	GED	-0.099	0.036	0.039	0.142	-0.026	0.408
	High-school graduate	-0.043	0.063	-0.019	0.269	-0.029	0.031
	Some college	-0.003	0.887	0.018	0.251	0.01	0.402
	College and above	Ref.		Ref.		Ref.	
Income-poverty ratio	<1	-0.052	0.161	-0.015	0.493	-0.028	0.208
	1-1.9	0.010	0.636	-0.028	0.073	-0.013	0.256
	2-3.9	-0.021	0.257	-0.012	0.416	-0.013	0.229
	4+	Ref.		Ref.		Ref.	
BMI (per 10 units)	BMI	0.046	$1.6x10^{-5}$	$3.0x10^{-4}$	0.973	0.018	0.006
Smoking	Never smoker	Ref.		Ref.		Ref.	
	Former smoker	-0.020	0.187	-0.05	$5.9x10^{-6}$	-0.037	$8.9x10^{-6}$
	Current smoker	-0.006	0.839	-0.082	$1.2x10^{-4}$	-0.05	0.004
Alcohol	0 drinks per week	Ref.		Ref.		Ref.	
	1-7 drinks per week	0.036	0.032	0.009	0.554	0.024	0.027
	7+ drinks per week	0.031	0.020	0.011	0.590	0.019	0.128
Daily activity	None	Ref.		Ref.		Ref.	
	Light	0.041	0.070	-0.036	0.030	-0.01	0.482
	Medium	0.057	0.003	-0.002	0.887	0.026	0.036
	Vigorous	0.045	0.053	0.027	0.266	0.034	0.033

[a]Sample sizes for male subjects Waves 1-9: n=1086, 1042, 993, 1030, 1035, 1042, 1053, 1059, 1086. Sample sizes for female subjects Waves 1-9: n=1270, 1229, 1190, 1235, 1226, 1216, 1228, 1241, 1270
[b]Estimates are from a linear regression of natural log of T/S ratio on smoking characteristics adjusted for age, race, education, income, body mass index, alcohol consumption, and daily physical activity.

Supplementary Table 3.6. Associations Between Telomere Length Measured at Wave 9 and Smoking-Related Parameters, Stratified by Sex, From the Health and Retirement Study, 2008

		Male				Female				All[a]			
		n	β[b]	P	P_{trend}	n	β[b]	P	P_{trend}	N	β[b]	P	P_{trend}
Pack-years (max packs/day*years smoked)[c]	Never	724	Ref.	0.253		1,661	Ref.	$5.7x10^{-7}$		2,385	Ref.	$1.4x10^{-5}$	
	<5	207	-0.007	0.735		270	-0.020	0.218		477	-0.015	0.225	
	5-<15	169	-0.035	0.114		214	-0.023	199		383	-0.029	0.034	
	≥15	643	-0.014	0.336		417	-0.070	$2.8x10^{-7}$		1060	-0.041	$2.9x10^{-5}$	
Pack-years (packs/day*years smoked)[d]	Never	724	Ref.	0.001		1,661	Ref.	$1.7x10^{-10}$		2,385	Ref.	$2.9x10^{-12}$	
	<5	174	-0.039	0.087		244	-0.072	$4.4x10^{-5}$		418	-0.060	$1.8x10^{-5}$	
	5-<15	141	-0.051	0.039		170	-0.086	$2.1x10^{-5}$		311	-0.070	$6.4x10^{-6}$	
	≥15	224	-0.062	0.002		243	-0.088	$3.5x10^{-7}$		467	-0.078	$2.8x10^{-9}$	

[a]Missing data among subjects with telomere length measurement was 61% (1925/3138) for pack-years smoked, and 62% (1941/3138) for max pack-years smoked. Telomere length was longer for those missing pack-years (logT/S ratio=0.24 vs 0.21, *P*=0.002), and not different by missing max pack-years (*P*=0.586)

[b]Estimates are from a linear regression of natural log of T/S ratio on smoking parameters adjusted for sex (when appropriate), age, race, education, income, body mass index, alcohol consumption, and daily physical activity

[c]Maximum estimate of pack-years from smoking duration in years multiplied by "cigarettes smoked per day when subject was smoking the most in their life", as answered by subjects on the questionnaire

[d]Lifetime pack-years based on "on-study" pack-year as described in main text, and "off-study" pack-year based on the "cigarettes smoked per day" upon study entry as an estimate for "off-study" smoking rate for smokers, and based on the "cigarettes smoked per day when subject was smoking the most in their life" for former smokers entering the study without any reported "on-study" smoking rate.

Supplementary Table 3.7. Associations between Telomere Length and Smoking in Simulated Datasets where Loss to Follow-up is Influenced by Smoking Status and Telomere Length

Effect on probability of loss to follow-up		**No effect (β = 0)**				**Negative effect (β = -0.200)**			
		All Participants		Participants not lost to follow-up		All Participants		Participants not lost to follow-up	
OR_{smoke}	OR_{TL}	median β	mean β	median β	mean β	median β	mean β	median β	mean β
1	1	0.001	0.000	0.001	0.001	-0.200	-0.200	-0.200	-0.199
1	0.75	0.001	0.001	0.001	0.001	-0.200	-0.200	-0.195	-0.195
1	0.5	0.001	0.000	-0.001	-0.002	-0.199	-0.200	-0.182	-0.181
1.5	1	-0.001	-0.001	-0.001	-0.002	-0.200	-0.200	-0.200	-0.200
1.5	0.75	-0.001	-0.001	0.027	0.027	-0.200	-0.200	-0.171	-0.172
1.5	0.5	0.003	0.002	0.056	0.058	-0.200	-0.200	-0.129	-0.130
2	1	-0.002	-0.002	-0.002	-0.002	-0.200	-0.199	-0.198	-0.198
2	0.75	0.000	-0.001	0.046	0.047	-0.199	-0.200	-0.150	-0.151
2	0.5	0.001	0.000	0.094	0.093	-0.202	-0.202	-0.089	-0.087

Each median/mean beta estimate is obtained from 1,000 simulated datasets, each consisting of 5,000 individuals (prior to excluding individuals lost to follow-up)

CHAPTER 4: A STUDY OF TELOMERE LENGTH, ARSENIC EXPOSURE, AND ARSENIC TOXICITY IN A BANGLADESHI COHORT

Authors: Chenan Zhang[1], Muhammad G. Kibriya[1], Farzana Jasmine[1], Shantanu Roy[1], Habibul Ahsan[1,2,3,4], Brandon L. Pierce[1,2,3,*]

Affiliations: [1]Department of Public Health Sciences, University of Chicago, Chicago IL 60615; [2]Department of Human Genetics, University of Chicago, Chicago IL 60615; [3]Comprehensive Cancer Center, University of Chicago, Chicago IL 60615; [4]Department of Medicine, University of Chicago, Chicago IL 60615

Funding: This work was supported by the National Institutes of Health [T32AG000243 and P30AG012857 to C.Z., and R01 ES020506 and U01 HG007601 to B.L.P.)

ABSTRACT

Chronic arsenic exposure has been shown in epidemiologic studies to increase risk for arsenical skin lesions and cancer. Recent studies suggest that the mechanism of arsenic's toxicity may be related to telomere dysfunction. However, prior studies of arsenic exposure, telomere length (TL), and skin lesion tend to be small and cross-sectional. We investigated the associations between arsenic exposure and TL, and between prospective TL and incident skin lesions among individuals participating in the Health Effects of Arsenic Longitudinal Study in Araihazar, Bangladesh. Quantitative PCR was used to measure the average TL of peripheral blood DNA samples collected at baseline. Multi-variable linear regression was used to estimate the association between baseline arsenic exposure (well water and urine) and TL in a random subcohort, and a nested case-control study was used to estimate the association between TL and incident skin lesion risk (diagnosed <8 years after baseline). We observed no association between baseline arsenic exposure and TL, and this was true for both water- and urine-based exposure measures (n=1,170). Comparing incident skin lesion cases with sex- and age-matched controls (448 cases, 462 controls), we observed higher skin lesion risk among individuals with shorter TL (P_{trend}=4.6x10^{-5}), with odds ratios of 3.05, 1.30, and 1.21 for the first (shortest), second, and third TL quartiles compared to the longest. In conclusion, arsenic exposure was not found to be associated with TL in this Bangladesh population, suggesting that leukocyte TL does not reflect a primary mode of action for arsenic's toxicity. However, short TL may be a biomarker of susceptibility for arsenical skin lesions independent of arsenic exposure.

.

INTRODUCTION

More than 100 million people worldwide experience chronic arsenic exposure through naturally contaminated drinking water (1), including approximately 20-45 million in Bangladesh (2). Arsenic exposure through drinking water has been reported to increase the risk of various adverse health outcomes including mortality (3), neurological conditions (4), cardiovascular diseases (5), as well as malignancies such as cancers of the skin, bladder, kidney, liver, and lung (6–10). An early characteristic sign of arsenic toxicity is the appearance of skin lesions, which can reflect susceptibility to arsenic-related disease and arsenic-related cancers (10–14). Arsenic-induced skin lesions can also be a source of social stigmatization and discrimination resulting in adverse social and mental health issues (15).

The mechanism of arsenic's toxicity and carcinogenicity may be related to telomere length and dysfunction. Telomeres are the repeating six nucleotide sequence binding protein complexes at the end of human chromosomes that protect DNA ends from damage. Telomere length (TL) shortening occurs with cell division, and has been investigated as a biomarker of aging and susceptibility to age-related health conditions, including cardiovascular diseases, neurocognitive diseases, cancers, and overall mortality (16–20), many of which have also been linked with chronic arsenic exposure. Experimental studies have shown that arsenic exposure can both increase and decrease TL and telomerase activity depending on dose (21,22). TL shortening has been attributed to oxidative stress and inflammation (23–25), two processes also identified as potential modes of chronic arsenic toxicity (26–28). Recent epidemiologic studies suggest that arsenic exposure increases TL in peripheral blood cells (29–31) and alters expression of genes involved in telomere maintenance (30–32). However, the few existing studies of the association between arsenic exposure and TL tend to be small, and only one study to date investigates the

association between TL and arsenical skin lesions—a cross-sectional study that found longer TL in skin lesion subjects compared to those without skin lesions (29). A prospective study assessing the association between baseline TL and risk of subsequent arsenic-related skin lesions is needed to establish the temporal relationship between TL and skin lesions, and allow for the evaluation of TL as a biomarker for predicting arsenic-related skin lesions in exposed populations.

In this study, we investigated the association between arsenic exposure and TL, and between prospectively measured TL and subsequent skin lesion among individuals with a wide range of arsenic exposure through drinking water in the Health Effects of Arsenic Longitudinal Study (HEALS) in Araihazar, Bangladesh.

METHODS

Study participants

HEALS is a prospective cohort study designed to assess the effects of exposure to arsenic through drinking water on health. Details of the participant selection, study design, and study methods have been described elsewhere (33). Briefly, 11,746 male and female participants were recruited between October 2000 and May 2002 in Araihazar, Bangladesh. Demographic and lifestyle data, as well as blood and urine samples were collected at baseline. In-person visits were conducted every two years following the baseline visit, and all visits included a physical examination and collection of urine. Informed consent was obtained from all study participants, and the study protocol was approved by the Institutional Review Boards of the University of Chicago, Columbia University, and the ethical committee of the Bangladesh Medical Research Council.

Skin lesion status

During the baseline interview, participants were examined by trained physicians for the presence of arsenical skin lesions according to a structured protocol. The physicians, who were blinded to the participants' well water arsenic exposure, recorded the presence or absence of melanosis (hyperpigmentation), leucomelanosis (hypopigmentation), or keratosis (thickening of the skin on palms and soles). For this study, skin lesion status was defined as the presence or absence of any type of the aforementioned skin lesions. Over the course of the study period of three biennial follow-ups after baseline, 866 individuals developed incident skin lesions out of the 10,182 individuals included at baseline (34).

Random subcohort selection:

A random subset of the baseline HEALS cohort was selected, including sampling of participants with prevalent skin lesions at baseline, subjects with incident skin lesions, and skin lesion controls (34,35). We chose to conduct the analysis of the association between arsenic and TL in a population-based subcohort rather than in a selected group of skin-lesion controls or skin-lesion cases in order to avoid potential bias arising from conditioning on a collider (36), i.e. biased estimates between arsenic and TL due to conditioning on skin lesion status, a potential common effect of both arsenic exposure and TL.

Nested case-control study control selection:

To assess associations between TL and skin lesion status, we conducted a nested case-control study consisting of 506 individuals with incident skin lesions who were frequency matched by sex and 5-year age intervals with 506 subjects who remained skin lesion-free during the study. Sampling was done with preference for subjects with DNA already extracted. We chose to perform age- and sex- matching for this analysis because age and sex are well-described correlates of TL (18), which may confound the association between arsenic and TL.

Exposure assessment

Baseline exposure to arsenic was determined by asking participants to identify the primary well used as their main source of drinking water. Water from 5,966 wells in the study area was analyzed for arsenic concentration by graphite furnace atomic absorption spectrometry with a detection limit of 5.0 μg/L. Samples below the limit of detection were reanalyzed by inductively coupled plasma-mass spectrometry with a detection limit of 0.1 μg/L (37). Spot urine samples were obtained at baseline in order to measure urinary total arsenic concentration by graphite furnace atomic absorption spectrometry with a detection limit of 2 μg/L (38), and urinary creatinine by a colorimetric Sigma Diagnostics Kit (Sigma, St. Louis, MO). Urinary arsenic was divided by creatinine concentration to generate a creatinine-adjusted urinary arsenic concentration (μg/g) (39). Daily arsenic intake (μg/day) was calculated by multiplying the primary well water arsenic concentration (μg/L) by the self-reported daily water volume consumed from that well (L/day), with secondary well information also incorporated in the calculation. Water arsenic exposure was modeled as quartiles with cut points for the first and second quartiles adjusted to correspond to WHO's guideline for arsenic in drinking water (≤10

μg/L), and Bangladesh's national standard for arsenic in drinking water (≤50 ug/L). Creatinine-adjusted urinary arsenic concentration and daily arsenic intake were categorized by quartiles based on the distribution of each measure in the random subcohort.

TL measurement

Genomic DNA was extracted from whole blood extracted from clot blood using Flexigene DNA kit (Cat # 51204) from Qiagen. Two quantitative PCR (qPCR) protocols with different plate designs were used in the measurement of TL (Plate Design 1 and 2), with telomere and 36B4 target sequence primers for reference and subject samples loaded onto a 96-well plate for each plate design as detailed in the Supplementary Materials and Supplementary Figures 4.1 and 4.2. Skin lesion cases and controls were randomized across plates, and lab technicians were blinded to the status. TL of the samples in the nested case-control study (n=920) and a subset of the samples in the subcohort study (n=460) were measured using Plate Design 1 and a protocol previously modified and validated by Ehrlenbach (40). In brief, qPCR was used to compare the abundance of the telomere repeat sequence to that of a single-copy gene (36B4) in a given sample, with both telomere and single-copy gene amplification of each DNA sample performed on the same plate. Reference samples were included on each plate to allow standardization of each reaction run with respect to inter-plate variations. The TL of the remaining samples in the subcohort study (n=1009) were measured using Plate Design 2 and a protocol previously described by Cawthon (41). In brief, telomere and single-copy gene amplification of each DNA sample were performed on paired plates, with the sample placed at the same position on each of the two plates.

Both plate designs were run with samples containing 30 μL amplification reaction volume of 1× Qiagen Quantitect Sybr Green Master Mix (Qiagen) and 25 ng of template DNA. The primer sequences for both plate designs were as follows: 5′GGTTTTTGAGGGTGAGGGTGAGGGTGAGGGTGAGGGT3′ (forward), 5′TCCCGACTATCCCTATCCCTATCCCTATCCCTATCCCTA3′ (reverse) for telomere sequence amplification and 5'CAGCAAGTGGGAAGGTGTAATCC3' (forward), 5'CCCATTCTATCATCAACGGGTACAA3' (reverse) for 36B4 sequence amplification. QPCRs for Plate Design 1 were performed on 96-well plates containing both primer sets with the following protocol: 95 °C for 15 min to activate Taq polymerase; 30 cycles at 95 °C for 15 s, 54°C for 2 min, and 72°C for 1 min; followed by 72°C for 4 min. QPCRs for Plate Design 2 involved two sets of PCR conditions. The telomere reaction was run with the following protocol: 95 °C for 15 min to activate Taq polymerase; 25 cycles at 95 °C for 15 s and 54°C for 2 min. The 36B4 reaction was run with the following protocol: 95 °C for 15 min to activate Taq polymerase; 30 cycles at 95 °C for 15 s and 58°C for 1 min.

C_t values (the number of PCR cycles needed for amplicons to cross a threshold of background fluorescence) were averaged across replicates for each sample, and T/S ratio was calculated using the following equation:

$$T/S = \frac{2^{Ct_{\text{telomere}}(Reference)}/2^{Ct_{telomere}(Sample)}}{2^{Ct_{36B4}(Reference)}/2^{Ct_{RPLOP}(Sample)}}$$

TL quartiles were created based on the distribution of TL in the random subcohort. We calculated the coefficients of variation (CV) for each set of replicates of a sample target within a plate using the C_t. CV was calculated as the standard deviation of the sample replicates divided by the mean of the replicates:

$$CV = 100 \times \sqrt{\frac{\sum(x - \bar{x})^2}{n}} \div \bar{x}$$

As a quality control step, we excluded up to one replicate within each triplicate measure of the sample (and up to two replicates within each 6 replicate measures of the reference sample) with outlying values that would cause the CV to exceed 1% for samples measuring 36B4, and 3% for samples measuring telomere sequence. If the CV still exceeded the set thresholds after exclusion of outlying value(s), all measurements for the participant were excluded. As a second quality control step, we pooled C_t values across all telomere sequence measurements and all 36B4 gene measurements, respectively, and excluded outlying C_t values from each if they exceeded three standard deviations from the mean. As a third quality control step, we calculated the T/S ratio for all remaining samples, and excluded a sample if it exceeded three standard deviations from the mean. Samples for 91 subjects were excluded from the matched case-control samples, with samples from 921 subjects (456 cases and 465 controls) used for subsequent analyses in the nested case-control study. Samples for 49 subjects were excluded, with samples from 1,469 subjects used for subsequent analyses in the subcohort study.

To assess the reproducibility of the assay for the first protocol, measurements of 37 randomly selected samples were repeated on different days, and a CV was calculated for the paired samples as the pooled standard deviation of the duplicate measures (x_1 and x_2) divided by the overall mean of all the measurements.

$$CV = 100 \times \sqrt{\frac{\sum(x_1 - x_2)^2}{2n}} \Bigg/ \frac{\sum(x_1 + x_2)}{2n}$$

The TL and single-copy gene assay CVs for triplicates, i.e. intra-plate CVs, were less than 3%. The paired CV for the 37 duplicate samples, i.e. inter-plate CV, was 11.7%. To assess the

reproducibility of the assay for the second protocol, an additional reference sample was measured across 31 plates, from which the calculated inter-plate CV was 9.8%.

Covariates

The following covariates, measured at baseline, were included as covariates in all analyses: age, sex, education (0, 1-5, 6-16 years), body mass index (BMI) (<18.5, 18.5-24.9, >25), smoking (never, former, current), TV and land ownership (measures of socioeconomic status), and high blood pressure (systolic blood pressure ≥ 140 mm Hg and/or diastolic blood pressure ≥ 90 mm Hg). Sun exposure was not assessed as a covariate in this study because it is uncommon among women in this study population to be exposed to the sun, and prior literature does not indicate an association between sun exposure and TL (42).

Statistical analyses

We assessed the association between arsenic exposure and TL in subjects from the random subcohort using linear regression (n = 1,469). To remove the variation in TL due to PCR plate position (intra-plate differences account for 5-9% of the variation in TL measure) or plate number (inter-plate differences account for 13-25% of the variation in TL measure), we first obtained the position- and plate-adjusted residual TL measure for each plate design using a mixed-effects model with continuous TL as the outcome, position as a fixed effect, and plate as a random effect. The position- and plate-adjusted residual TL was then used as the outcome in minimally-adjusted linear regression models with TL as the outcome and arsenic exposure as the predictor, accounting for continuous age, sex, plate design type, and DNA extraction batch (a study design variable), while fully-adjusted models additionally included socioeconomic factors,

BMI, and high blood pressure as covariates. Analyses were repeated using water arsenic, urinary arsenic, and daily arsenic dose (modeled as quartiles) as the arsenic exposure of interest.

Next, we calculated the odds ratios (ORs) for the risk of skin lesions by quartile TL using conditional logistic regression accounting for matched sets based on sex and age. We again adjusted for plate position and plate number, along with age, sex, and DNA extraction batch in minimally-adjusted conditional logistic regression models, and included socioeconomic factors, BMI, high blood pressure, and arsenic exposure as covariates in fully-adjusted models. The analyses of arsenic association with TL, and TL association with skin lesion were repeated after stratifying by sex and median age (<=45, >45), and tests of multiplicative interaction were performed by evaluating the Wald statistic for the cross-product interaction term from the regression model. We also investigated whether arsenic was an effect modifier of the association between TL and skin lesion with tests of multiplicative interaction assessing the *P* value of the cross-product term of the ordinal exposure variable (TL) and the binary effect modifier (high/low arsenic exposure) in the conditional logistic regression.

All analyses were done with Stata (version 14, StataCorp, College Station, TX).

RESULTS

Subject characteristics

Figure 1 shows a flowchart of the selection of participants for the nested case-control study. After TL measurement quality control measures were implemented, TL measures for 921 subjects for the nested case control study remained for subsequent analyses. The baseline characteristics, arsenic exposure, and TL measures of the age- and sex- matched skin lesion cases (n=456) and controls (n=464) used for the nested case-control study are shown in Table 4.1,

Figure 4.1. Selection of Samples for Nested Case-Control, Health Effects of Arsenic Longitudinal Study, Bangladesh, 2000-2009.

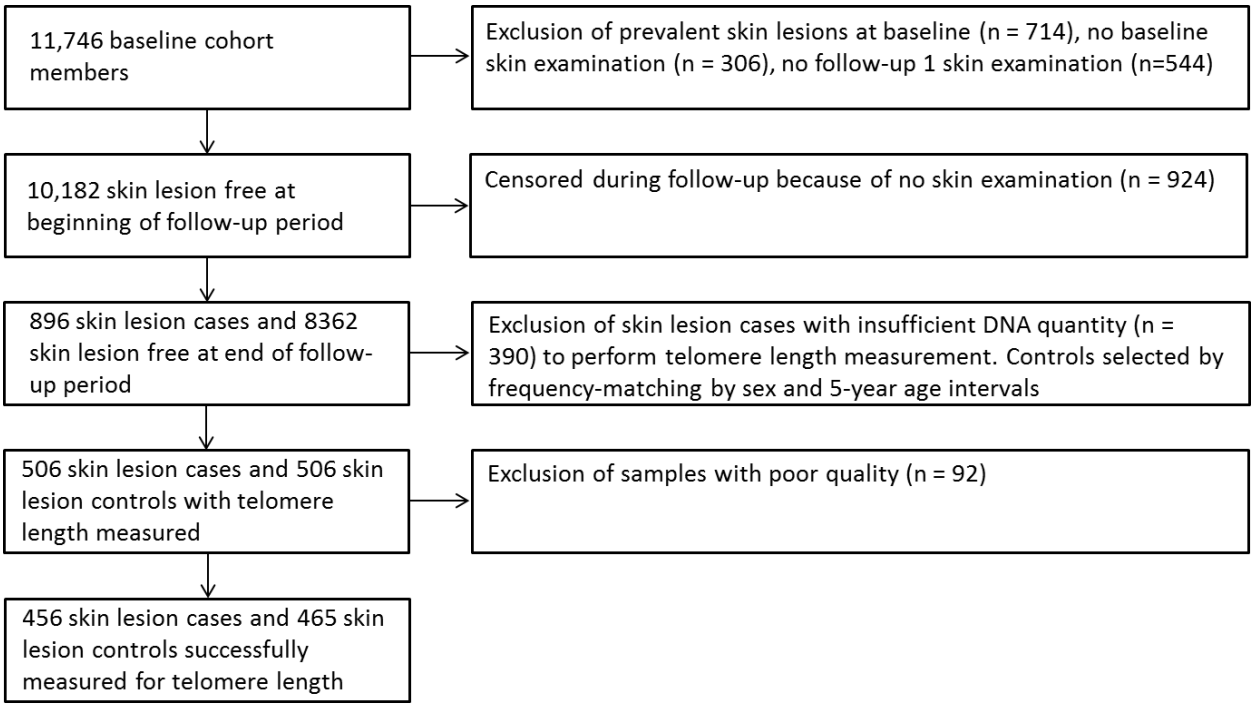


Table 4.1. Baseline Characteristics of incident skin lesions cases, matched controls, and a random sample of the HEALS cohort, Health Effects of Arsenic Longitudinal Study, Bangladesh, 2000-2009[a]

		Skin lesion cases	Matched skin lesion controls	Random subcohort
n		456	464	1469
Sex	Male	316 (69.3%)	314 (67.7%)	541 (36.8%)
	Female	140 (30.7%)	150 (32.3%)	928 (63.2%)
Age	years	44.6 (9.2)	44.8 (9.1)	35.1 (9.1)
Education	0	223 (49.0%)	191 (41.2%)	603 (41.0%)
	1-5	124 (27.3%)	138 (29.7%)	465 (31.7%)
	6-16	108 (23.7%)	135 (29.1%)	401 (27.3%)
BMI	<18.5	190 (42.1%)	191 (41.2%)	525 (35.9%)
	18.5-24.9	241 (53.4%)	239 (51.5%)	822 (56.2%)
	>25	20 (4.4%)	34 (7.3%)	116 (7.9%)
Smoking	Never	184 (40.4%)	198 (42.6%)	1014 (69.0%)
	Former	68 (14.9%)	57 (12.3%)	67 (4.6%)
	Current	204 (44.7%)	209 (45.0%)	388 (26.4%)
Own TV	Yes	144 (31.6%)	183 (39.4%)	505 (34.4%)
Own land	Yes	261 (57.2%)	277 (59.7%)	758 (51.6%)
Hypertension	Yes	76 (16.7%)	79 (17.1%)	201 (13.8%)
Water Arsenic (ug/L)	0.1-10	70 (15.4%)	148 (31.9%)	362 (24.6%)
	10.1-50	58 (12.7%)	111 (23.9%)	342 (23.3%)
	50.1-150	131 (28.7%)	132 (28.4%)	449 (30.6%)
	150.1+	197 (43.2%)	73 (15.7%)	316 (21.5%)
Creatinine-adjusted Urinary Arsenic (ug/g)	<105	81 (18.4%)	143 (31.0%)	366 (24.9%)
	105-<194	81 (18.4%)	138 (29.9%)	358 (24.4%)
	194-<329	112 (25.4%)	96 (20.8%)	363 (24.7%)
	329+	167 (37.9%)	85 (18.4%)	382 (26.0%)
Daily Arsenic dose (ug /day)	<30.4	69 (15.2%)	147 (31.7%)	366 (24.9%)
	30.4-<147	73 (16.1%)	125 (26.9%)	363 (24.7%)
	147-<376	123 (27.1%)	92 (19.8%)	368 (25.1%)
	376+	189 (41.6%)	100 (21.6%)	372 (25.3%)
T/S quartiles	<0.6897	160 (35.1%)	111 (23.9%)	368 (25.1%)
	0.6897-<0.7835	106 (23.2%)	110 (23.7%)	367 (25.0%)
	0.7835-0.9015	109 (23.9%)	127 (27.4%)	367 (25.0%)
	0.9015+	81 (17.8%)	116 (25.0%)	367 (25.0%)

[a]Categorical variables are shown in counts and percentages; continuous variables in mean± standard deviations (SD).

along with characteristics of the random subcohort (n=1,469) used for the arsenic-TL association analysis. TL was significantly inversely associated with age within the random subcohort (P=1.1x10^{-14}) (Supplementary Figure 4.3).

Association between arsenic and TL

Arsenic exposure was not associated with TL in the random subcohortfor any of the three exposure measures, with null estimates for both minimally-adjusted models accounting for only age, sex, and design variables, and when additionally adjusting for sociodemographic and lifestyle characteristics (Table 4.2). We observed no modification of the association between arsenic exposure and TL by sex or age (not shown).

Association between TL and skin lesion

Short TL was associated with increased skin lesion incidence in a dose-dependent manner, with similar estimates for both minimally-adjusted models accounting for only age, sex, and design variables (P_{trend}=2.3x10^{-5}), and when additionally adjusting for sociodemographic and lifestyle characteristics (P_{trend}=1.4x10^{-5}) (Table 4.3). Adjustment for arsenic exposure type did not appreciably change the estimates across the three measures of exposure (results not shown). Stratified analyses by skin lesion subtype (melanosis, leucomelanosis, or keratosis) resulted in similar estimates (results not shown).

Table 4.2. Association between Baseline Arsenic Exposure and Telomere Length, Health Effects of Arsenic Longitudinal Study, Bangladesh, 2000-2009 (n=1469)

	Minimally-adjusted model[a]			Covariate-adjusted model[b]		
Water Arsenic (ug/L)	β	CI	P	β	CI	P
.1-10	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
10.1-50	-0.017	-0.037,0.003	0.098	-0.018	-0.038,0.002	0.079
50.1-150	0.006	-0.012,0.025	0.508	0.004	-0.015,0.023	0.686
150.1+	-0.001	-0.021,0.020	0.958	-0.002	-0.023,0.018	0.844
	P_{trend} = 0.473			P_{trend} = 0.597		
Urinary Arsenic (ug/g)						
<105	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
105-<194	0.000	-0.020,0.019	0.979	-0.001	-0.021,0.019	0.937
194-<329	-0.004	-0.024,0.015	0.681	-0.006	-0.025,0.014	0.568
329+	0.008	-0.011,0.028	0.393	0.005	-0.014,0.025	0.585
	P_{trend} = 0.481			P_{trend} = 0.707		
Daily arsenic dose (ug/day)						
<30.4	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
30.4-<147	-0.008	-0.027,0.012	0.441	-0.009	-0.029,0.011	0.372
147-<376	0.006	-0.014,0.025	0.572	0.003	-0.017,0.022	0.793
376+	0.011	-0.009,0.030	0.279	0.009	-0.010,0.029	0.351
	P_{trend}=0.146			P_{trend}=0.209		

[a]Adjusting for age, sex, and laboratory batch variables
[b]Additionally adjusting for BMI, education, smoking status, TV ownership, land ownership, and high blood pressure

Table 4.3. Odds Ratios for Incident Skin Lesions According to Quartiles of Baseline Telomere Length, Health Effects of Arsenic Longitudinal Study, Bangladesh, 2000–2009

Model	Telomere length	Case/ Control	OR	95% CI	P
Minimally-adjusted[a]	4 (longest)	78/116	1.00	Ref.	Ref.
	3	98/116	1.12	0.68,1.83	0.66
	2	120/116	1.63	1.00,2.66	0.05
	1	160/116	2.61	1.59,4.26	$1.4x10^{-4}$
	Total	456/464		P_{trend} =$2.3x10^{-5}$	
Fully-adjusted[b]	4 (longest)	76/115	1.00	Ref.	Ref.
	3	94/116	1.30	0.78,2.19	0.31
	2	116/115	1.72	1.03,2.87	0.04
	1	160/116	2.96	1.76,4.97	$4.1x10^{-5}$
	Total	446/462		P_{trend} =$1.4x10^{-5}$	

[a]Adjusting for age, sex, and laboratory batch variables
[b]Additionally adjusting for BMI, education, smoking status, TV ownership, land ownership, high blood pressure, and daily arsenic dose

Interaction analyses

We observed no modification of the association between TL and skin lesion by age. In the stratified analysis of the association between TL and skin lesion by sex, estimates were larger for women than men, i.e. short TL was associated with larger increased risk in skin lesions among female subjects. However, the *P* value of the cross-product term of TL and being female not significant, potentially due to the smaller sample size of women in the study (Supplementary Table 4.1). We also observed significant interactions with arsenic exposure for the association between TL and skin lesion risk for all three measures of arsenic exposure ($P_{interactions}$=0.01-0.05), with higher risk among subjects exposed to lower arsenic levels (Table 4.4).

Table 4.4. Arsenic-Stratified Odds Ratios for Incident Skin Lesions According to Quartiles of Baseline Telomere Length, Health Effects of Arsenic Longitudinal Study, Bangladesh, 2000–2009

		Low Arsenic			High Arsenic			
Telomere length	Arsenic measure	Case/ Control	OR	95% CI	Case/ Control	OR	95% CI	$P_{interaction}$
4 (longest)	Water Arsenic	13/65	1.00	Ref.	65/50	1.00	Ref.	
3		35/72	1.73	0.69,4.34	59/44	1.11	0.55,2.24	
2		27/67	2.41	0.92,6.32	89/48	1.75	0.89,3.46	
1		48/54	5.99	2.30,15.6	112/62	2.15	1.09,4.25	0.05
4 (longest)	Urinary Arsenic	18/72	1.00	Ref.	60/43	1.00	Ref.	
3		36/72	1.61	0.75,3.45	58/44	0.92	0.43,1.98	
2		36/70	2.22	0.99,4.99	80/45	1.38	0.67,2.87	
1		66/66	5.46	2.45,12.2	94/50	1.72	0.83,3.57	0.01
4 (longest)	Daily arsenic dose	15/66	1.00	Ref.	63/49	1.00	Ref.	
3		39/76	1.60	0.69,3.72	55/40	1.10	0.52,2.31	
2		29/70	2.06	0.85,4.99	87/45	1.74	0.88,3.46	
1		54/59	5.16	2.17,12.3	106/57	2.48	1.23,5.01	0.05

DISCUSSION

In this large cohort-based study of Bangladeshi men and women, we found no association between baseline arsenic exposure and baseline TL; however, we observed a significant association between short baseline TL and increased risk of subsequent arsenical skin lesions. In addition, we found a significant interaction between TL and arsenic exposure, with a stronger association between short TL and skin lesion risk observed at lower levels arsenic exposure compared to high levels.

Our observation of a lack of association between arsenic exposure and TL differed from prior epidemiologic studies, which suggested that arsenic exposure increases TL (31,29,30). However, it is important to note that all three prior studies were limited by small sample sizes (n=60-200). In addition, the study by Chatterjee et al. observed no difference in TL when

comparing lesion-free subjects with high arsenic exposure to lesion-free subjects with low exposure. Rather, Chatterjee et al. only observed a positive association between high arsenic exposure and longer TL among subjects with skin lesions, an association potentially susceptible to collider bias due to stratifying on skin lesion status, a potential common outcome of both arsenic exposure and TL (36). Additionally, studies by both Chatterjee et al and Li et al were limited to comparing TL in high arsenic exposed subjects from one region to low arsenic exposed subjects from a different region – a design susceptible to potential biases due to confounding by other local factors.

Our observation of increased skin lesion risk associated with short TL was in contrast to the only other epidemiologic study on TL and arsenical skin lesions (29), which found increased skin lesion risk for individuals with longer TL. Previous *in vitro* studies of arsenic exposure on cell cultures suggest a bimodal effect of arsenic on TL in which low doses either maintain or lengthen TL, while high doses shorten TL (21,22,28). However, it is unclear if the range of arsenic concentrations tested in vitro in these studies corresponds to the same range of exposures experienced by human hematopoietic cells *in vivo* via arsenic-contaminated drinking water. Furthermore, *in vitro* studies performed over hours, days, or even weeks may not be ideal for extrapolating effects corresponding to years of chronic arsenic exposure.

There is no prior literature with which to compare the effect modification we observed of the stronger association between short TL and skin lesion at lower compared to higher arsenic exposure levels. However, a previous study describing an association between skin lesion risk and metabolites of inorganic arsenic only present at a lower arsenic exposure level (<50 μg/L) (43) suggests that some risk factors of skin lesion are conditional on the level of arsenic exposure. This is in alignment with more general findings of the paradoxical effects of arsenic in

which both carcinogenic and anticancer effects have been attributed to arsenic exposure (44,45), with *in vitro* studies showing protective effects against oxidative stress and DNA damage at low arsenic levels and cytotoxic effects at higher levels (46,47). The attenuation of the TL-skin lesion association at higher arsenic exposures may be another example of such a phenomenon, although the underlying mechanism of this heterogeneity is unclear. Additional investigations to replicate and elucidate this finding are necessary.

Although TL was not associated with arsenic exposure in our study, TL was found to be an independent predictor of arsenical skin lesions. As a biomarker of skin lesion risk, telomere length may serve as a valuable predictor susceptibility to arsenic toxicity, and furthermore provide opportunities to better understand characteristics of susceptible individuals who are in need for intervention and prevention of arsenical skin lesion.

This study has several key strengths: for the arsenic-TL association, we used a random subset from a cohort study, which is more representative of the population of interest, thus reducing the risk of selection bias; both analyses had much larger sample sizes than prior studies, providing superior statistical power; our analyses involved more comprehensive adjustment of potential confounders as covariates; and we were able to assess a dose-response association for the TL-skin lesion association. Our study of TL and skin lesion risk is prospective in design, with controls selected from the same cohort as the cases, reducing potential issues of reverse causality and selection bias. This is the first study to our knowledge in which TL was measured prior to skin lesion ascertainment. Furthermore, this study had approximately 95% or greater follow-up among eligible participants at each follow up period, with subjects experiencing a wide range of arsenic exposures, and multiple types of arsenic exposure measurements collected.

One potential limitation to this work is the need to assume that TL measures from blood samples approximate that of the skin – our primary target of interest. This assumption is likely valid in normal (low arsenic) conditions as prior findings indicate that there is high correlation in TL between an individual's leukocyte and skin (r=0.83) (48). However, not all tissues in the body accumulate arsenic equally, with studies in both human and mice models showing that skin tends to accumulate higher levels of arsenic than blood (due in part to arsenic's affinity to the higher levels of keratin in skin) (49–51). It is possible that chronic and high arsenic exposure could be associated with TL in skin cells (or other tissues), but not the blood TL measures analyzed in this study. Future analyses examining arsenic levels and TL from biopsied skin samples may be necessary to clarify this issue.

Another potential limitation of the arsenic-TL association analysis is that measures of arsenic are not representative of a cumulative lifetime exposure to arsenic. However, participants selected into the HEALS study were required to be residents of the study area for at least 5 years and primarily drinking water from one of the study wells for at least 3 years (33), so we believe the estimates of arsenic exposure at baseline to be representative of stable long-term exposure through drinking water. Nevertheless, it is worth examining in future analyses the impact of different exposure windows on the association between arsenic and TL, particularly in children and/or adolescents, who may represent a more susceptible group to chronic arsenic exposure. Furthermore, longitudinal studies of arsenic exposure and TL would allow better ascertainment of the association between lifetime arsenic exposure and TL, as well as clarify the best time to measure TL as a predictive biomarker for subsequent health outcomes.

CONCLUSIONS

Our findings suggest a lack of a strong association between arsenic exposure and blood TL. However, our observation that short TL is a prospective biomarker for arsenical skin lesions has important implications for predicting individual susceptibility to arsenic-induced skin lesions and potentially to other arsenic-related health conditions. The differing impact of the level of arsenic exposure on this association requires further research to elucidate potential biological pathways responsible in order to develop and evaluate effective interventions.

REFERENCES

1. World Health Organization. United Nations Systhesis Report on Arsenic in Drinking Water. Geneva: WHO; 2001.

2. Flanagan SV, Johnston RB, Yan Zhenga. Arsenic in tube well water in Bangladesh: health and economic impacts and implications for arsenic mitigation. Bull World Health Organ. 2012;90:839–846 8p.

3. Sohel N, Persson LÅ, Rahman M, Streatfield PK, Yunus M, Ekström E-C, et al. Arsenic in Drinking Water and Adult Mortality: A Population-based Cohort Study in Rural Bangladesh. Epidemiology. 2009;20:824–30.

4. Vahidnia A, van der Voet GB, de Wolff FA. Arsenic neurotoxicity — A review. Hum Exp Toxicol. 2007;26:823–32.

5. States JC, Srivastava S, Chen Y, Barchowsky A. Arsenic and Cardiovascular Disease. Toxicol Sci. 2009;107:312–23.

6. Celik I, Gallicchio L, Boyd K, Lam TK, Matanoski G, Tao X, et al. Arsenic in drinking water and lung cancer: a systematic review. Environ Res. 2008;108:48–55.

7. Hopenhayn-Rich C, Biggs ML, Smith AH. Lung and kidney cancer mortality associated with arsenic in drinking water in Córdoba, Argentina. Int J Epidemiol. 1998;27:561–9.

8. Liu J, Waalkes MP. Liver is a Target of Arsenic Carcinogenesis. Toxicol Sci. 2008;105:24–32.

9. Mink PJ, Alexander DD, Barraj LM, Kelsh MA, Tsuji JS. Low-level arsenic exposure in drinking water and bladder cancer: a review and meta-analysis. Regul Toxicol Pharmacol. 2008;52:299–310.

10. Yu H-S, Liao W-T, Chai C-Y. Arsenic Carcinogenesis in the Skin. J Biomed Sci. 2006;13:657–66.

11. Cuzick J, Evans S, Gillman M, Price Evans DA. Medicinal arsenic and internal malignancies. Br J Cancer. 1982;45:904–11.

12. Cuzick J, Sasieni P, Evans S. Ingested arsenic, keratoses, and bladder cancer. Am J Epidemiol. 1992;136:417–21.

13. Hsu L-I, Chen G-S, Lee C-H, Yang T-Y, Chen Y-H, Wang Y-H, et al. Use of Arsenic-Induced Palmoplantar Hyperkeratosis and Skin Cancers to Predict Risk of Subsequent Internal Malignancy. Am J Epidemiol [Internet]. 2013; Available from: http://aje.oxfordjournals.org/content/early/2013/01/07/aje.kws369.abstract

14. Arsenic in Drinking Water: 2001 Update. Washington (DC): National Academies Press (US); 2001.

15. Brinkel J, Khan MH, Kraemer A. A Systematic Review of Arsenic Exposure and Its Social and Mental Health Effects with Special Reference to Bangladesh. Int J Environ Res Public Health. 2009;6:1609–19.

16. Cawthon RM, Smith KR, O'Brien E, Sivatchenko A, Kerber RA. Association between telomere length in blood and mortality in people aged 60 years or older. The Lancet. 2003;361:393–5.

17. Martin-Ruiz C, Dickinson HO, Keys B, Rowan E, Kenny RA, Von Zglinicki T. Telomere length predicts poststroke mortality, dementia, and cognitive decline. Ann Neurol. 2006;60:174–80.

18. Sanders JL, Newman AB. Telomere Length in Epidemiology: A Biomarker of Aging, Age-Related Disease, Both, or Neither? Epidemiol Rev. 2013;35:112–31.

19. Serrano AL, Andrés V. Telomeres and Cardiovascular Disease: Does Size Matter? Circ Res. 2004;94:575–84.

20. Willeit P, Willeit J, Mayr A, Weger S, Oberhollenzer F, Brandstätter A, et al. Telomere length and risk of incident cancer and cancer mortality. J Am Med Assoc. 2010;304:69–75.

21. Ferrario D, Collotta A, Carfi M, Bowe G, Vahter M, Hartung T, et al. Arsenic induces telomerase expression and maintains telomere length in human cord blood cells. Toxicology. 2009;260:132–41.

22. Zhang T-C, Schmitt MT, Mumford JL. Effects of arsenic on telomerase and telomeres in relation to cell proliferation and apoptosis in human keratinocytes and leukemia cells in vitro. Carcinogenesis. 2003;24:1811–7.

23. Jenny NS. Inflammation in aging: cause, effect, or both? Discov Med. 2012;13:451–60.

24. O'Donovan A, Pantell MS, Puterman E, Dhabhar FS, Blackburn EH, Yaffe K, et al. Cumulative Inflammatory Load Is Associated with Short Leukocyte Telomere Length in the Health, Aging and Body Composition Study. PLoS ONE. 2011;6:e19687.

25. Von Zglinicki T. Oxidative stress shortens telomeres. Trends Biochem Sci. 2002;27:339–344.

26. Ahmed S, Mahabbat-e Khoda S, Rekha RS, Gardner RM, Ameer SS, Moore S, et al. Arsenic-associated oxidative stress, inflammation, and immune disruption in human placenta and cord blood. Environ Health Perspect. 2011;119:258–64.

27. Kitchin KT. Recent advances in arsenic carcinogenesis: modes of action, animal model systems, and methylated arsenic metabolites. Toxicol Appl Pharmacol. 2001;172:249–61.

28. Liu L, Trimarchi JR, Navarro P, Blasco MA, Keefe DL. Oxidative Stress Contributes to Arsenic-induced Telomere Attrition, Chromosome Instability, and Apoptosis. J Biol Chem. 2003;278:31998–2004.

29. Chatterjee D, Bhattacharjee P, Sau TJ, Das JK, Sarma N, Bandyopadhyay AK, et al. Arsenic exposure through drinking water leads to senescence and alteration of telomere length in humans: A case-control study in West Bengal, India. Mol Carcinog. 2015;54:800–9.

30. Gao J, Roy S, Tong L, Argos M, Jasmine F, Rahaman R, et al. Arsenic exposure, telomere length, and expression of telomere-related genes among Bangladeshi individuals. Environ Res. 2015;136:462–9.

31. Li H, Engström K, Vahter M, Broberg K. Arsenic Exposure through Drinking Water Is Associated with Longer Telomeres in Peripheral Blood. Chem Res Toxicol. 2012;25:2333–9.

32. Mo J, Xia Y, Ning Z, Wade TJ, Mumford JL. Elevated human telomerase reverse transcriptase gene expression in blood cells associated with chronic arsenic exposure in Inner Mongolia, China. Environ Health Perspect. 2009;117:354–60.

33. Ahsan H, Chen Y, Parvez F, Argos M, Hussain AI, Momotaj H, et al. Health Effects of Arsenic Longitudinal Study (HEALS): Description of a multidisciplinary epidemiologic investigation. J Expos Sci Env Epidemiol. 2005;16:191–205.

34. Argos M, Kalra T, Pierce BL, Chen Y, Parvez F, Islam T, et al. A Prospective Study of Arsenic Exposure From Drinking Water and Incidence of Skin Lesions in Bangladesh. Am J Epidemiol. 2011;174:185–94.

35. Ahsan H, Chen Y, Parvez F, Zablotska L, Argos M, Hussain I, et al. Arsenic Exposure from Drinking Water and Risk of Premalignant Skin Lesions in Bangladesh: Baseline Results from the Health Effects of Arsenic Longitudinal Study. Am J Epidemiol. 2006;163:1138–48.

36. Cole SR, Platt RW, Schisterman EF, Chu H, Westreich D, Richardson D, et al. Illustrating bias due to conditioning on a collider. Int J Epidemiol. 2010;39:417–20.

37. Cheng Z, Zheng Y, Mortlock R, van Geen A. Rapid multi-element analysis of groundwater by high-resolution inductively coupled plasma mass spectrometry. Anal Bioanal Chem. 2004;379:512–8.

38. Nixon DE, Mussmann GV, Eckdahl SJ, Moyer TP. Total arsenic in urine: palladium-persulfate vs nickel as a matrix modifier for graphite furnace atomic absorption spectrophotometry. Clin Chem. 1991;37:1575–9.

39. Nermell B, Lindberg A-L, Rahman M, Berglund M, Persson LÅ, El Arifeen S, et al. Urinary arsenic concentration adjustment factors and malnutrition. Environ Res. 2008;106:212–8.

40. Ehrlenbach S, Willeit P, Kiechl S, Willeit J, Reindl M, Schanda K, et al. Influences on the reduction of relative telomere length over 10 years in the population-based Bruneck Study: introduction of a well-controlled high-throughput assay. Int J Epidemiol. 2009;38:1725–34.

41. Cawthon RM. Telomere measurement by quantitative PCR. Nucleic Acids Res. 2002;30:e47–e47.

42. Sugimoto M, Yamashita R, Ueda M. Telomere length of the skin in association with chronological aging and photoaging. J Dermatol Sci. 2006;43:43–7.

43. Jansen RJ, Argos M, Tong L, Li J, Rakibuz-Zaman M, Islam MT, et al. Determinants and consequences of arsenic metabolism efficiency among 4,794 individuals: demographics, lifestyle, genetics, and toxicity. Cancer Epidemiol Biomarkers Prev [Internet]. 2015; Available from: http://cebp.aacrjournals.org/content/early/2015/12/16/1055-9965.EPI-15-0718.abstract

44. Smith AH, Marshall G, Yuan Y, Steinmaus C, Liaw J, Smith MT, et al. Rapid Reduction in Breast Cancer Mortality With Inorganic Arsenic in Drinking Water. EBioMedicine. 2014;1:58–63.

45. Waxman S, Anderson KC. History of the Development of Arsenic Derivatives in Cancer Therapy. The Oncologist. 2001;6:3–10.

46. Snow ET, Sykora P, Durham TR, Klein CB. Arsenic, mode of action at biologically plausible low doses: What are the implications for low dose cancer risk? Living Safe Chem World Proc 10th Int Congr Toxicol 11-15 July 2004 Tamp Finl. 2005;207:557–64.

47. Yang P, He X-Q, Peng L, Li A-P, Wang X-R, Zhou J-W, et al. The role of oxidative stress in hormesis induced by sodium arsenite in human embryo lung fibroblast (HELF) cellular proliferation model. J Toxicol Environ Health A. 2007;70:976–83.

48. 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.

49. Hughes MF, Kenyon EM, Edwards BC, Mitchell CT, Razo LMD, Thomas DJ. Accumulation and metabolism of arsenic in mice after repeated oral administration of arsenate. Toxicol Appl Pharmacol. 2003;191:202–10.

50. Hughes MF, Thomas DJ, Kenyon EM. Toxicology and Epidemiology of Arsenic and its Compounds. Arsenic [Internet]. John Wiley & Sons, Ltd; 2009. page 237–75. Available from: http://dx.doi.org/10.1002/9780470741122.ch4

51. Saha JC, Dikshit AK, Bandyopadhyay M, Saha KC. A Review of Arsenic Poisoning and its Effects on Human Health. Crit Rev Environ Sci Technol. 1999;29:281–313.

SUPPLEMENTARY MATERIALS

Relative TL of the samples in the nested case-control study and a subset of the samples in the subcohort study was measured using a protocol previously modified and validated by Ehrlenbach (1). Each plate consisted of triplicates of 14 participant DNA samples and six replicates of a reference DNA sample, each measured for both telomere and single copy gene for a total of 96 samples per plate. The 14 possible positions and corresponding primers for each plate using this design are described in Supplementary Figure 4.1.

Relative TL of the remaining samples in the subcohort study was measured using a protocol previously described by Cawthon (2). Each plate consisted of triplicates of 31 participant DNA samples and triplicates of a reference DNA sample, each measured for either telomere or single copy gene at the same position on paired plates. The 31 possible positions for each plate using this design are described in Supplementary Figure 4.2.

1. Ehrlenbach S, Willeit P, Kiechl S, Willeit J, Reindl M, Schanda K, et al. Influences on the reduction of relative telomere length over 10 years in the population-based Bruneck Study: introduction of a well-controlled high-throughput assay. Int J Epidemiol. 2009;38:1725–34.

2. Cawthon RM. Telomere measurement by quantitative PCR. Nucleic Acids Res. 2002;30:e47–e47.

Supplementary Figure 4.1. Layout of quantitative PCR plate consisting of triplicates of 14 participant DNA samples and six replicates of a reference DNA sample, each measured for telomere (light gray cells) and single copy gene (dark gray cells) for a total of 96 samples per plate.

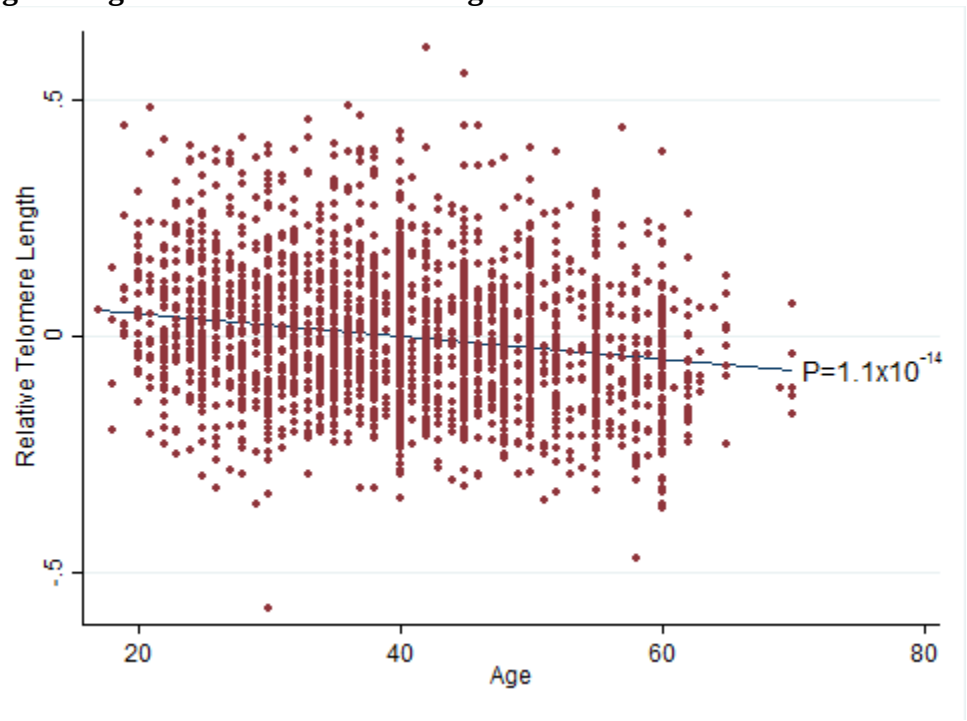
	1	2	3	4	5	6	7	8	9	10	11	12
A	1						2					
B	3						4					
C	5						6					
D	7						8					
E	9						10					
F	11						12					
G	13						14					
H	ref						ref					

Supplementary Figure 4.2. Layout of quantitative PCR plate consisting of triplicates of 31 participant DNA samples and triplicates of a reference DNA sample, each measured for either telomere or single copy gene at the same position on paired plates. Light gray represents the first plate consisting of telomere primers only, dark gray represents the second plate consisting of single copy gene primers only.

	1	2	3	4	5	6	7	8	9	10	11	12
A	1			2			3			4		
B	5			6			7			8		
C	9			10			11			12		
D	13			ref			14			15		
E	16			17			18			19		
F	20			21			22			23		
G	24			25			26			27		
H	28			29			30			31		

	1	2	3	4	5	6	7	8	9	10	11	12
A	1			2			3			4		
B	5			6			7			8		
C	9			10			11			12		
D	13			ref			14			15		
E	16			17			18			19		
F	20			21			22			23		
G	24			25			26			27		
H	28			29			30			31		

Supplementary Figure 4.3. Scatterplot of baseline relative leukocyte telomere length against age overlaid with a fitted regression line.



Supplementary Table 4.1. Sex-Stratified Odds Ratios for Incident Skin Lesions According to Quartiles of Baseline Telomere Length, Health Effects of Arsenic Longitudinal Study, Bangladesh, 2000–2009

		Men			Women			
Telomere length	Arsenic measure	Case/ Control	OR	95% CI	Case/ Control	OR	95% CI	$P_{interaction}$
4 (longest)	Water Arsenic	48/57	1.00	Ref.	31/50	1.00	Ref.	
3		68/86	1.25	0.66,2.37	34/41	4.74	0.73,30.7	
2		70/79	1.47	0.78,2.79	30/30	11.3	1.99,64.6	
1		126/91	2.95	1.55,5.61	41/28	7.7	1.22,48.4	0.60
4 (longest)	Urinary Arsenic	48/56	1.00	Ref.	29/50	1.00	Ref.	
3		66/86	0.91	0.48,1.72	32/41	3.01	0.50,18.0	
2		69/78	1.34	0.72,2.52	30/30	11.8	2.16,64.4	
1		122/91	2.54	1.35,4.78	38/28	10.4	1.66,65.6	0.52
4 (longest)	Daily arsenic dose	48/57	1.00	Ref.	30/50	1.00	Ref.	
3		67/86	1.18	0.63,2.23	34/41	3.19	0.51,19.84	
2		70/79	1.44	0.77,2.69	30/30	9.03	1.71,47.83	
1		126/91	2.90	1.54,5.47	41/28	9.50	1.51,59.63	0.49

CHAPTER 5: CONCLUSIONS

SUMMARY OF RESULTS

In Chapter 2 I assessed the associations between nine telomere length (TL)-associated SNPs and risk for five common cancer types (breast, lung, colorectal, ovarian and prostate cancer, including subtypes) using a Mendelian randomization approach. The genetic score for long TL was significantly associated with an increased risk of lung adenocarcinoma, but not another lung cancer subtype, squamous cell carcinoma. Furthermore, TL was not associated with any other cancer type investigated, suggesting that findings from prior retrospective measures of TL may be biased. In Chapter 3 I characterized the association between TL and non-genetic determinants, including demographic, socioeconomic, and lifestyle factors, with an emphasis on smoking. Smoking was associated with shorter TL when using prospective measures of smoking among men and women, but the association was strongly attenuated in the cross-sectional analysis among men. This attenuation was attributed to a higher rate of smoking cessation during the study period among males with shorter TL compared to males with longer TL. Short TL was also associated with poorer overall health in men, suggesting that male smokers with short TL were more likely to quit smoking due to poor health. In Chapter 4 I investigated whether TL was cross-sectionally associated with chronic arsenic exposure, and prospectively associated with subsequent skin lesion risk in a Bangladeshi cohort. Short TL was not found to be associated with chronic arsenic exposure, but was prospectively associated with increased subsequent skin lesion risk in a Bangladeshi cohort.

SYNTHESIS OF FINDINGS

Findings from the three studies in this dissertation may initially appear to be incongruent with one another. These include the observations that 1) while long TL is a risk factor for lung cancer (a disease with smoking as a well-known risk factor), smoking itself appears to be a risk factor for short TL; 2) while arsenic and TL are not associated, short TL appears to be a risk factor for arsenical skin lesions; and 3) while long TL is a risk factor for lung cancer, it appears to be protective against skin lesions. However, I suggest that these findings are complementary and reflect TL's multiple potential roles as a biomarker of effect, a biomarker of susceptibility, and a cause of cancer.

The role of TL as a causal contributor to cancer reflects the hypothesized mechanism that long telomeres increase the proliferative duration of cells, thus delaying senescence and allowing further mutations to occur (or conversely, that short telomeres trigger cell senescence or programmed cell death, protecting against the onset of carcinogenesis). This characteristic of TL is highlighted by the findings of the Mendelian randomization study in Chapter 2, and represents an inherited genetic trait of the individual. The role of TL as a biomarker of effect (i.e., the effect of an exposure) reflects the impact of environmental exposures on TL (whether to lengthen or shorten it), with those environmental exposures (rather than TL itself) being the causal agents contributing to carcinogenesis. This characteristic of TL is highlighted by the findings of the Health and Retirement Study in Chapter 3, and represents the cumulative effects of environmental exposures on an individual's TL. Finally, the role of TL as a biomarker of susceptibility reflects the ability of an individual to respond to the challenge of an environmental

exposure that may lead to cancer. This characteristic of TL is highlighted by the findings of the Bangladesh Cohort Study in Chapter 5, and represents an inherent quality of the individual that is independent of environmental exposures.

In Chapters 2 and 3, the observations that long TL is a risk factor for lung cancer while smoking is a risk factor for short TL may initially appear paradoxical. Because the association between long TL and lung cancer was inferred from genetic evidence (which is not susceptible to environmental exposures such as smoking history), it is unlikely that the Chapter 2 finding is explained by the quitting phenomenon observed in Chapter 3. However, in context of the multiple roles of TL as a biomarker, the finding from Chapter 2 (long TL as a risk factor for lung cancer) primarily reflects TL's role as a cause of cancer, while the finding from Chapter 3 (smoking is a risk factor for short TL) primarily reflects TL's role as a biomarker of smoking effect. Furthermore, it must be emphasized here that separate investigations of the associations between smoking and TL, and TL and lung cancer, do not suffice as methods to establish TL as a mediator of the association between smoking and cancer. In fact, evidence from these two chapters would suggest that TL is clearly not a primary mediator of the association between smoking and cancer. Conversely, taken together, these findings strengthen the argument that TL plays dual roles as both a causal contributor to cancer risk (through potential biological effect of long TL on increased cancer risk) as well as a biomarker of exposure to carcinogens present in tobacco smoke (with short TL being a proxy for the effect of smoking on increased cancer risk), with both TL and tobacco making opposing causal contributions to cancer risk. Prior literature suggests that smoking has a much weaker association with adenocarcinoma than squamous cell

carcinoma and other subtypes of lung cancer (relative risk = 3.1 vs 13.5) (1). Should smoking exist as a residual confounder in TL-lung cancer analysis, this heterogeneity potentially explains why an association was observed between long TL and lung adenocarcinoma (where the opposing causal contribution of smoking is weaker), but not squamous cell carcinoma, (where the opposing causal contribution of smoking is stronger and could mask the causal contribution of TL).

In Chapter 4, observations that arsenic and TL are not associated, but short TL appears to be a risk factor for arsenical skin lesions may initially appear inconsistent. However, we can view the lack of association between arsenic and TL as a reflection of TL's role as a biomarker of effect (or lack thereof for arsenic in this analysis), while TL as a risk factor for arsenical skin lesion may primarily reflect TL's role as a biomarker of susceptibility. Similar to the explanation addressing the "paradoxical" findings between Chapters 2 and 3, separate observations of the associations between arsenic and TL, and TL and skin lesion suggest that TL should not be viewed as a mediator of the association between arsenic exposure and skin lesion.

In Chapters 2 and 4, the observations that long TL is a risk factor for lung cancer while also being associated with decreased skin lesion risk may appear contradictory. However, we can distinguish the finding of long TL as a risk factor for lung cancer as a reflection of TL's hypothesized role as a cause of lung adenocarcinoma, while the finding of long TL being protective against skin lesions may primarily reflect that of TL's role as a biomarker of susceptibility to the effects of arsenic.

Finally, it is important to note that while the various potential roles of TL as different types of biomarkers can exist independently of each other, it is likely that a combination of two or all three together are relevant to an individual's overall susceptibility to cancer. While a causal inference method such as Mendelian randomization may be helpful for parsing the role of TL as a causal contributor to cancer independent of the other two components, other traditional epidemiology studies may not allow such distinction. Therefore, when using observational study designs to investigate the association between TL and cancer, it is necessary to recognize the nuanced and multi-faceted role TL plays as a potential biomarker of cancer, particularly in context of causal inference.

STRENGTHS AND LIMITATIONS

A primary strength of this dissertation is the diverse types of study designs used to characterize TL determinants, and elucidate the associations between TL and cancer. This included the use Mendelian randomization, a type of instrumental variable analysis originating from econometrics causal inference literature; cross-sectional and prospective analyses of smoking as a TL determinant; and prospective analysis of TL as a risk factor for skin lesion. This multi-dataset, multi-method approach is a critical step towards clarifying the causal relationship between TL and cancer. Additional strengths of individual studies include the large sample sizes of cancer cases and controls assessed for association with TL SNPs in Chapter 2, the rich covariate data allowing for comprehensive adjustments for demographic, socioeconomic, and lifestyle variables in a nationally representative cohort of older subjects in Chapter 3, and the

largest known sample size of TL analyses in a Bangladeshi cohort, a unique study population, in Chapter 4.

Limitations of individual studies were discussed in detail in Chapters 2, 3 and 4. In brief, the validity of the Mendelian randomization assumptions in Chapter 2 cannot be proven—we could only provide additional sensitivity analyses to demonstrate that the results were robust under various exclusion criteria of SNPs included in the SNP score. Furthermore, our power to detect associations may have been limited due to the small variance in measured TL explained by SNPs used in the analysis (1-2%) (2). However, power analyses indicated that the detection of strong to moderate associations was possible with the large sample sizes provided by the dataset. The detection of weak associations require even larger sample sizes, or the inclusion of additional SNPs in the SNP score should more telomere-associated SNPs be discovered in future studies. Finally, TL measured from DNA samples collected from subjects may vary depending on the tissue source of the DNA, e.g. saliva, blood, skin. In the study described in Chapter 3, this may not be an issue since there is no specific tissue target of interest for health outcome assessment, and therefore no *a priori* reason to favor any particular DNA source. However, in Chapter 4, we needed to assume that TL measures from blood samples approximate that of the skin – our primary target of interest. While prior studies showing high correlation in TL between an individual's leukocyte and skin (3) suggest that the assumption is likely valid, this assumption needs to be further confirmed by measuring TL in banked biopsied skin tumors and adjacent normal skin, which is available for the majority of participants in the Bangladesh cohort study showing signs of skin cancer.

SIGNIFICANCE AND IMPACT

This dissertation project has significant impact on current knowledge, future research, and public health implications. Parsing the genetic and environmental factors associated with TL has two benefits. Should genetic proxies of TL prove to be associated with cancer risk, future cancer prediction efforts can utilize a TL multi-SNP risk score as a novel risk measure tool. Analysis of environmental factors leads to better understanding of the modifiability of TL as a potential risk factor for cancer, allowing the future development of targeted intervention of telomere attrition through lifestyle or dietary strategies. Furthermore, future cancer prevention efforts could consider targeting TL as a pharmacological target, particularly for those individuals determined to be at higher risk for telomere-length associated cancer. Addressing these knowledge gaps is a critical step towards clarifying the causal relationship between TL and cancer, with the long-term goal of using this knowledge to improve cancer prediction and prevention.

REFERENCES

1. Jedrychowski W, Becher H, Wahrendorf J, Basa-Cierpialek Z, Gomola K. Effect of tobacco smoking on various histological types of lung cancer. J Cancer Res Clin Oncol. 1992;118:276–82.

2. Pierce BL, Ahsan H, Vanderweele TJ. Power and instrument strength requirements for Mendelian randomization studies using multiple genetic variants. Int J Epidemiol. 2011;40:740–52.

3. 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.