Genetic variation in telomere maintenance genes and survival after colorectal cancer diagnosis

Hang Yin

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Science

University of Washington

2018

Committee:

Polly A. Newcomb

Li Hsu

Sara Lindstroem

Program Authorized to Offer Degree:

Epidemiology

University of Washington

Abstract

Genetic variation in telomere maintenance genes and survival after colorectal cancer diagnosis

Hang Yin

Chair of the Supervisory Committee:

Polly A. Newcomb

Department of Epidemiology

Introduction: Telomeres are the repetitive nucleotide sequences capping the end of chromosomes and are regulated by telomerase. Both telomere length and telomerase initiate carcinogenesis and prognosis in patients diagnosed with cancer, including colorectal cancer (CRC). Genome-wide association studies (GWAS) have identified several loci related to telomere length or telomerase activity, but the association between those single nucleotide polymorphisms (SNPs) and survival after CRC diagnosis remains uncertain. Smoking and sex influence telomere/telomerase, so we also investigate the gene-smoking and gene-sex interaction with CRC survival.

Methods: We conducted large gene-wide study with 4896 invasive CRC cases from Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO). Common variants within 13 genes

involved in telomere maintenance (*TERT, TERC, TERF1, TERF2, TINF2, TERF2IP, ACD, POT1, TNKS, TNKS2, TNKS1BP1, TEP1* and *PINX1)* were included. Multivariate Cox proportional hazard models were fit to analyze the association between these variants and overall and CRC-specific survival, with additional stratification analyses according to smoking status, smoking pack-years and sex. Likelihood ratio tests were used to test the significance of interaction terms. P-values were adjusted for multiple comparison by Bonferroni method.

Results: Several SNPs within *TERT, TERF1, TNKS, TNKS1BP1, TEP1* and *TERF2* were associated with overall and/or CRC-specific survival at the p-value threshold of 0.05. After Bonferroni adjustment, the association between rs7200950 (*ACD***)** and CRC-specific survival was significantly modified by categorical smoking pack-years (adjusted P=0.049 for interaction). The minor allele at rs74429678 (*POT1*) increased CRC-specific mortality in women (HR: 1.33, 95%CI: 1.07-1.65) but not in men (HR: 0.75, 95%CI: 0.52-0.97); the minor allele at rs2975843 (*TERF1*) increased overall mortality in women (HR: 1.08, 95%CI: 0.99-1.18) but not in men (HR: 0.84, 95%CI: 0.75-0.92). Gene-wide significant interaction was also detected between sex and rs75676021 (*POT1*, adjusted P=0.023 for interaction).

Conclusion: Our study reported the gene-wide statistically significant interaction between genes involved in telomere maintenance and smoking pack-years (*ACD*) and sex (*POT1, TERF1*), respectively. Our study also suggested an association between candidate genes and overall/CRC-specific survival. Validation of these findings by other large studies and further functional annotation on those SNPs may be warranted.

Introduction

Colorectal cancer (CRC) is the second leading cause of cancer-related deaths in men and women in the United States (1). Improved treatment, and early detection have both contributed to overall increased survival. Despite this, CRC remains a major cause of morbidity and mortality throughout the world (2). Therefore, understanding more about prognostic factors may provide potential approaches to CRC control.

Telomeres are relevant to human diseases (3). They cap the ends of eukaryotic chromosomes and are comprised of repetitive nucleotide sequences (4). Telomeres act by protecting the ends of chromosomes from deterioration or from end-to-end fusion with neighboring chromosomes (5), thus preventing aberrant chromosomal replication and help maintain chromosomal stability and genomic integrity. This telomere replication is regulated by telomerase complex, which is made up of telomerase reverse transcriptase (encoded by *TERT*) and RNA component (encoded by *TERC*) , shelterin complex (encoded by *TERF1, TERF2, TINF2, TERF2IP, ACD,* and *POT1*) (6,7) and other associated proteins (encoded by *TNKS, TNKS2, TNKS1BP1,TEP1* and *PINX1*) (8). Telomeres change over time, shortening with each cell division, partly due to incomplete replication of the 3'-end of the chromosomes (4). Age, sex and environmental factors like cigarette smoking may also impact telomere function (9,10). Dysfunction in telomere replication mechanisms result in accelerated genetic changes and cellular senescence. Hence, telomeres are considered to be a hallmark of aging.

Telomeres and telomerase together play an important role in CRC carcinogenesis through the well-characterized chromosomal instability (CIN) pathway (11,12). Functional variants in

telomere maintenance genes may affect telomerase expression and further influence cancer risk and progression through their effect on telomerase activity and telomere length. Genome-wide association studies (GWAS) have identified that the 5p15.33 genomic region containing *TERT* and cleft lip and palate transmembrane 1-like (*CLPTM1L*) genes are associated with multiple cancers, including lung, breast, bladder, prostate, cervical and CRC (13–15).Variants within the *TERC* and *POT1* region have been associated with CRC risk (16,17). The relationship of these germline genetic variants with CRC survival is not yet well characterized.

Telomerase upregulation allows for cells to proliferate without limits. To support their immortal growth, most human cancer cells express high-levels of telomerase (18), and it may also serve as an important determinant of cancer prognosis. Indeed, genetic variation in telomere maintenance genes have been associated with overall and cancer-specific survival in cancers of the lung, glioma, liver, ovaries and breast (19–23). Furthermore, telomere/telomerase-based treatment brought benefits on increasing longevity and cancer therapy in mice models and human tumor cells and clinical trials are in process (24,25). Thus, drugs targeting on telomere/telomerases may serve as a viable option in cancer treatment.

No published studies have investigated whether telomere maintenance genes are specifically associated with survival after CRC diagnosis. To evaluate this association, we utilized data from the Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO) (26) to elucidate the relationship between single nucleotide polymorphism (SNP) variation in 13 telomere maintenance genes (TERT, TERC, *TERF1, TERF2, TINF2, TERF2IP, ACD, POT1, TNKS, TNKS2, TNKS1BP1, TEP1* and *PINX1*) and overall as well as CRC-specific survival after CRC

diagnosis. We also considered whether such associations are modified by smoking and sex, the demonstrated factors in telomere erosion.

Methods

Study Participants:

Study participants were drawn from 11 case-control and nested case-control studies, including study subjects from Colon Cancer Family Registry (Colon CFR) from US, Canada and Australia. In this study, we used data from six cohort studies in US cohorts: Health Professionals Follow-up Study (HPFS), Physicians' Health Study (PHS), VITamins And Lifestyle study (VITAL); Women's Health Initiative (WHI); Nurses' Health Study (NHS) and Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (PLCO); and two case-control studies from US, Canada and Europe: Diet, Activity and Lifestyle Study (DALS) and Postmenopausal Hormone study-Colon Cancer Family Registry (PMH-CCFR). Studies were describe in details previously (26).

Study subjects were restricted to participants self-reported as European descent with incident invasive CRC and whose genotype and survival information were available. CRC diagnosis was confirmed by medical record and pathology report. The outcome was all-cause death and CRC-specific death. Active follow-up was used to ascertain vital status in HPFS, PHS, NHS, PLCO, WHI, and dates and causes of deaths were confirmed by death certificate and/or medical record. For VITAL, DALS and PMH-CCFR, vital status was confirmed through National Death Index, cancer registries, state death records, or population registries with cause of death verified by death certificate. All participants gave informed consent and studies were approved by their Institutional Review Board (IRB) respectively. Characteristics of included studies are described in **Table 1**.

Data collection:

All demographic and clinical characteristics data were collected through self-report using questionnaires and/or in-person interview, and details haven been described previously (26). Data elements considered into analysis were age at diagnosis, sex, study center, cancer site, stage, smoking status, cigarettes smoking, pack-years and age stopped smoking. Survival related data included all-cause death, CRC-specific death, time from diagnosis to death or last follow-up. Principal components analysis (PCA) was used for population stratification. Details of genotyping, quality assurance and quality control (QA/QC) and imputation are described in supplementary material. Genomic DNA was extracted from blood sample or buccal cells by conventional methods. Genotyping platform of each studies was summarized in **Table 1**. Before imputation, genotyped SNPs were excluded based on call rate (< 98%), lack of Hardy-Weinberg Equilibrium in controls (HWE, $P < 1x10^{-4}$), and minor allele frequency (MAF < 5% for WHI Set 1, DALS Set 1; MAF < 5 / # of samples for each other study). All autosomal SNPs were imputed. Imputation was performed in HapMap II release 24 with MaCH, with reference population choosen from Utah residents with Northern and Western European Ancestry form the Center d'etude du polymorphisme humain (CEPH) collection (CEU). Based on a literature search, we included genes encoding proteins that participate in telomere length regulation. Genetic variation in some of the included genes have previously been associated with risk of CRC (13–17) and other cancer types as well as with cancer survival (19–23). Details on selected genes can be found in **Supplementary Table S1**. A total of 6578 SNPs was included, but only common (MAF≥5%) SNPs reached the sufficient power (details can be found in supplementary material). Finally, 1871 common SNPs were included in the analysis.

Statistical analysis:

Data were pooled from studies for statistical analysis. Within each gene, Pearson's correlation test was performed to check the correlation between every pairs of SNPs and principal components analysis (PCA) was performed to obtain the effect number (M_{eff_G}) of independent tests (27). M_{eff_G} was used for type I error control in Bonferroni correction in single-SNP model in survival analysis. Multivariate Cox proportional hazard regression was used to estimate hazard ration (HR) and 95% confidence intervals (95% CI) for the associations between single SNPs and survival. Separate model was constructed for overall and CRC-specific death. Schoenfeld residuals was examined to check the proportional hazards assumption. Log-additive models were used to assign values based on the copies of count allele (0=0 copies of count allele, 1=1 copies of count allele, 2=2 copies of count allele) for directly genotyped data. For imputed data, the dosage that represented the estimated number of copies of the count allele was provided. Dosage scaled from 0 to 2. All models were adjusted for age at diagnosis, sex (except for sex-interaction analysis), study center and the first three principal components of genetic ancestry. Adjusted p-values were calculated using M_{eff_G} of each gene as the number of independent tests and Bonferroni method. All estimates were from single-SNP models and the adjusted p-value <0.05 was considered as significance. Cigarette smoking history (yes/no), smoking status at the time of completing the questionnaire (former smoker/ current smoker/ never), categorical smoking pack-years (<12, 12-24, 25-44, ≥ 45) were environmental factors of interest in the GxE interaction. Sex was also included in the GxE test. Likelihood ratio tests were applied to evaluate whether the differences were significant. For the associations and/or interaction terms that reached the p-value <0.05, we chose the SNPs with the smallest p-value as the representative SNPs in the region of gene with high linkage disequilibrium (LD). LD was measured by Pearson correlation.

Results

This study included a total of 4896 invasive CRC cases. Among these cases there were 1098 all-cause deaths and 1681 CRC-specific deaths. The demographic and clinical characteristics of participants are summarized in **Table 2**. The majority of participants were women (62.4%) and aged 65 years and older (72.5%). A history of cigarette smoking was common among participants (56.3%), but only 18% were current smokers.

The association between the selected SNPs and overall and CRC-specific survival is presented in **Table 3**. At uncorrected P< 0.05, we identified several SNPs located at *TERT*, *TERF1*, *TNKS*, *TNKS1BP1*, *TEP1* and *TERF2*. Favorable overall and CRC-specific survival was associated with the minor allele at rs2075785 (*TERT*) and rs2981096 (*TERF1*). Increased overall mortality was associated with the minor allele at rs10102030 (*TNKS*), rs4939134 (*TNKS1BP1*) and rs1760894 (*TEP1*). The increased CRC-specific mortality was associated with rs251796 (TERF2) and rs10088969 (*TNKS*). None of the associations above remained significant after gene-level multiple comparison using Bonferroni adjustment.

Genetic associations with survival were evaluated in interactions with smoking (**Table 4**). The minor allele (G-allele) at rs35656875 (*TNKS*) was associated with favorable overall survival (HR: 0.70, 95%CI: 0.49-0.99) and CRC-specific survival (HR: 0.63, 95%CI: 0.41-0.98) among non-smokers, and was not associated with CRC-specific survival among smokers (HR: 1.04, 95%CI: 0.95-1.13). Neither interaction remained significant after multiple test adjustment. In the stratification analysis of grouped smoking pack-years, with P-value < 0.05, rs251797 (*TERF2*) and rs67456872 (*TNKS*) showed a statistically significant interaction with categorical smoking

pack-years in the association with both overall and CRC-specific survival (**Table 5**). Also, the association between rs7200950 (*ACD*) and CRC-specific survival was significantly modified by the categorical smoking pack-years after multiple comparison adjustment (adjusted P=0.049 for interaction). Furthermore, we detected that the minor allele (G-allele) at rs1865493 (*TERF2IP*) (adjusted P = 0.045 for interaction) was associated with favorable overall survival among smokers reporting ≥45 pack-years (HR: 0.56, 95%CI: 0.36-0.77), and this association was not be observed among non-smokers (HR: 0.98, 95%CI: 0.81-1.19) ; the minor allele (C-allele) at rs73202875 (*TNKS*) (adjusted *P* = 0.024 for interaction) was associated with increased CRC-specific mortality among non-smokers (HR: 1.37, 95%CI: 1.07-1.75) , while favorable CRC-specific survival among smokers reporting more than or equal to 45 pack-years (HR: 0.46, 95%CI: 0.19-0.74) (**Table 6**). There was not statistically significant interaction between genetic variants and current smoking status (never/current/former) (**Supplementary Table S2**).

We evaluated the selected genetic loci and association with survival according to sex (**Table 7**). Two SNPs within *POT1* (rs75676021, rs74429678) showed a gene-wide significant interaction with sex on the association with overall survival (adjusted P=0.023 for interaction) and CRC-specific survival (adjusted P=0.019 for interaction), respectively. The association between the minor allele (A-allele) at rs74429678 (*POT1*) and CRC-specific survival was poor in women (HR: 1.33, 95%CI: 1.07-1.65) but favorable in men (HR: 0.75, 95%CI: 0.52-0.97). Such difference between men and women was also observed in the correlation between minor allele (A-allele) at rs75676021 (*POT1*) and overall survival. Another SNP within *TERF1* (rs2975843) showed a gene-wide significant interaction with sex on the association with overall survival (adjusted P=0.002 for interaction) and CRC-specific survival (adjusted P=0.004 for interaction),

respectively. Women with the minor allele (A-allele) at rs2975843 (*TERF1*) had poor overall survival (HR: 1.08, 95%CI: 0.99-1.18), while inverse association was found in men (HR: 0.84, 95%CI: 0.75-0.92). Such difference on sex was observed in CRC-specific survival. The association between those SNPs and overall and CRC-specific survival is favorable in males but negative in females.

Discussion

In our large gene-wide study including 4896 colorectal cancer patients, we found several SNPs within *TERT, TERF1, TERF2, TEP1, TNKS* and *TNKS1BP1* that were associated with overall and/or CRC-specific survival after CRC diagnosis at the level of $P<0.05$. Such associations were not modified by smoking status (ever/never, current/former/never) at gene-wide significance level. Variants located at *ACD, TERF2IP, TNKS, POT1* and *TERF1* remained statistically differentially associated with survival according to smoking pack-years and sex.

Although no study reports appeared to have examined the telomere maintenance genes and CRC survival, some previous studies reported that some variants are associated with CRC susceptibility and survival in other cancer types. SNPs within *TERT (*rs2736100, rs2736098) (14,17,28), and *TERC* (rs10936599) (16) increased CRC susceptibility, but we did not detect a statistically significant association between these SNPs and survival. More variants within *TERT, TERF2, PINX1* and *TNKS* increased the overall mortality in patients with several cancers, include glioblastoma, bladder, lung, breast and ovaries (23,29–32). However, those results did not reach significance after multiple comparison correction and we also did not find these SNPs to be significant in our study. Other studies were moderately sized, which also limited their

power to investigate any interaction with smoking and sex as we did. Our results are consistent with a genome-wide study on the association between common variants and CRC prognosis. No SNPs included in our study were identified as genome-wide significantly associated with survival using discovery-based approach (33).

The dual roles of telomeres and telomerase are important in carcinogenesis and progression (34). Telomere length may initiate carcinogenesis through chromosomal instability, while it prohibits cell proliferation by inducing cell deaths; telomerase controls genetic instability and may also facilitate tumor cell growth. Telomere length in colorectal cancer cells was observed to be shorter than in normal mucosa (35), and telomere length in peripheral blood leukocyte appears to be an independent prognostic factor (35–38). Several in vitro and in vivo studies demonstrate association between high levels of telomerase/TERT and poor survival (39–41). Telomerase also participates in gene expression regulation, NF-κB signaling, cell growth and migration, which indicate it acts as a tumor-promoting factor, apparently independent of the elongation of telomere length (42–44). *TERT* gene transcription is the key determinant of the telomerase activity regulation, and a positive correlation was built on *TERT* mRNA and telomerase activity (45). Taken together, it is biologically plausible to expect the association between telomerase coding genes and colorectal cancer prognosis, but the evidence from the population study supporting this hypothesis was not strong.

There is considerable evidence that smoking plays an important role in telomere function, Cigarette smoking appeared to accelerate telomere length shortening by oxidative stress (46) and methylation (47,48). Long telomere length is inversely associated with smoking (49) and packs

per day among current smokers (50), Sung et al. found loci of *TNKS* and *PINX1* were associated with blood pressure using gene-smoking interaction model, but the interaction term failed to attain genome-wide significance (51). rs7200950 (*ACD*) and rs1865493 (*TNKS*) found by our study were intron variants, and rs73202875 (*TERF2IP*) is a promoter in 5'untranslated region (5'-UTR). *ACD* and *TERF2IP* are the shelterin complex subunit (6,52) and *TNKS* encodes poly (ADP-ribose) polymerases. Those genes are indispensable for telomere length regulation, *TERF2IP* regulates gene expression independently of regulation telomere length (53). Cumulative effects of short telomere length and smoking increased the risk of CRC, gastric cancer (GC) or esophageal squamous cell carcinoma (ESCC) in a Chinese Han population (54,55). Previous studies reported the negative correlation between telomerase activity/TERT expression and smoking status (56), although one study found inverse results in lung cancer (57), it provides an explanation for more favorable survival observed on smokers,. Furthermore, telomerase activity in non-small cell lung cancer (NSCLC) tumor cells was significantly lower in patients smoking less than 40 pack years comparing (58) to patients smoking more than 40 pack-years, this is consistent with our results that smokers with more than 45 pack-years had decreased risk of CRC-specific mortality in comparison to non-smokers in addition that high telomerase activity worsened survival outcome. So far, several polymorphisms in genes including DNA repair pathway, metabolic enzyme -smoking interaction on the association with colorectal cancer risk (59–61), but none of these variants related to telomerase and telomere and study investigating survival outcome or smoking pack-years is lacking.

Our study found a statistically significant association between rs75676021(A-allele, *POT1*), rs74429678(A-allele, *POT1*) and rs2975843 (A-allele, *TERF1*) with significant interaction with

sex increased overall and CRC-specific mortality in women but decreased both mortality in men. The results can be explained by the regulation of sex hormones on telomerase activity. Estrogen activates telomerase via up-regulating telomerase catalytic subunit or activating c-Myc/Max that binding to *TERT* promoter to increase its activity (62). Telomere length has also been found shorter in men compared to that in similarly aged women (63,64). Thus, our results are biologically plausible. In our study population, the percentage of non-smokers among women was higher than men (49% vs. 40%), and 20% of men were smokers with more than 45 pack-years compared to 10% of women only. Taking together, our finding in smokers with high dosage pack-years have smaller effect size than non-smokers; smoking, especially smoking dosage, may influence the gene-sex interaction.

To the best of our knowledge, this is the first study investigating the association between genetic variants involved in telomere maintenance and survival after colorectal cancer diagnosis. Our study has several strengths. First, the large samples size with long follow-up and complete and validated survival outcomes data facilitated the gene-wide main effect and GxE interaction evaluation. Second, we described smoking status by including recency and dosage to increase sensitivity. Third, our findings were supported by previous epidemiological studies on other cancer types and plausible biologic mechanisms. We also acknowledge some limitations. One is that we only included common variants in our analyses because low-frequency SNPs require larger sample sizes to reach detect true associations, so, we may have missed the fraction of survival outcome explained by low-frequency and rare variants. More powerful statistical methods are required to analysis those low-frequent variants. Also, we applied a candidate genes approach instead of genome-wide analysis because we were interested in genes related to

telomere maintenance function. We tried our best to include all genes of interest in our study, but it is possible that we missed additional genes contributing to telomere length regulation and our study may have had limited genetic diversity. Further, all autosomal SNPs were imputed and we used the expected number of copies of the minor allele. However, we restricted SNPs with high imputation accuracy in quality control procedure and previous report showed the imputed SNPs provided unbiased inference.

In conclusion, our candidate analysis in a large gene-wide study observed suggestive associations between genetic variation related to telomere maintenance function and overall and CRC-specific survival. We also detected the interaction between gene involved in telomere maintenance and smoking pack-years (*ACD, TERF2IP, TNKS*) and sex (*POT1, TERF1*) on the association with survival after colorectal cancer diagnosis. Current results need be verified by larger studies; further functional annotation about those statistically significant GxE variants and stratification analysis on both smoking status and sex may be of interest.

Acknowledgements

I would like to thank all people who make the study possible. For Harvard cohorts, I would like to thank the participants and staff of the HPFS, NHS, and PHS 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. For PLCO, I would like to thank the PLCO Cancer Screening Trial screening center investigators and the staff from Information Management Services Inc and Westat Inc. Most importantly, we thank the study participants for their contributions that made this study possible. For PMH, I would like to thank the study participants and staff of the Hormones and Colon Cancer study. For WHI, 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

I would like to thank with greatest gratitude my thesis committee chair, Dr. Polly A. Newcomb. I sincerely appreciated her expertise, patience, devotion and encouragement that help me complete my thesis with further publication. Also, I would like to thank my committee member, Dr. Sara Lindstroem and Dr. Li Hsu, for their expertise in genetic epidemiology and biostatistics. Their helpful and valuable comments contribute a lot to my thesis and stretch my thinking in my work. I would like to thank Dr. Sheetal Hardikar. She contributed a lot in study design and results interpretation with her expertise in the area of telomeres in relation to colorectal cancer. I would

like to thank Dr. Ulrike Peters for her support on my work and Barbara Banbury for her hard work to ensure the clean data with high quality as well as her help in addressing my problems. In addition, I would like to thanks all investigators, staff and researchers in GECCO and Newcomb's study group for their experience and advice on my study.

I would like to thank UW SPH for the two-years wonderful graduation study. The systematic training in epidemiology, advanced statistical analysis and programming skills I got definitely contribute to my further education or career. I am grateful for my cohort of fellow students and my friends. They made my life colorful and memorable. Thanks for their company and encouragement along the way. Finally, I want to give special thanks to my parents for their love, understanding and support for my dreams in pursuing a master degree in US.

Thank you all!

Funding

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

Harvard cohorts (HPFS, NHS, PHS): HPFS is supported by the National Institutes of Health (P01 CA055075, UM1 CA167552, U01 CA167552, R01 CA137178, R01 CA151993, R35 CA197735, K07 CA190673, and P50 CA127003), NHS by the National Institutes of Health (R01 CA137178, P01 CA087969, UM1 CA186107, R01 CA151993, R35 CA197735, K07 CA190673, and P50 CA127003) and PHS by the National Institutes of Health (R01 CA042182).

Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO): National Cancer Institute, National Institutes of Health, U.S. Department of Health and Human Services (U01 CA164930, R01 CA059045). This research was funded in part through the NIH/NCI Cancer Center Support Grant P30 CA015704.

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. Funding 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.

PMH: National Institutes of Health (R01 CA076366 to P.A. Newcomb).

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

Tables

Table 1. Characteristics of included studies

Study	Case (N)	Male/Female (N /N)	Age (years)[a]	N. deaths, all-cause [b]	N. deaths, CRC [c]	Median follow-up time (days)	Platform*
DALS 1	710	403/307	65.1 (32-79)	244 (34.4%)	135 (55.3%)	1917.56	610K, 550K
DALS 2	415	220/195	65.1 (31-79)	115 (27.7%)	81 (70.4%)	1674.1	300K
HPFS	168	168/0	71.5 (50-90)	82 (48.8%)	47 (57.3%)	2007.5	730K
NHS	296	0/296	68.1 (46-85)	118 (39.9%)	89 (75.4%)	2296.45	730K
PHS	324	324/0	70.6 (44-92)	200 (61.7%)	131 (65.5%)	2062.25	730K
PLCO 1	531	301/230	68.9 (55-82)	180 (33.9%)	108 (60.0%)	2430	300/240S, 610K
PLCO 2	478	275/203	70.4 (55-86)	103 (21.6%)	75 (72.8%)	1237.5	300K
PMH	279	0/279	64.4 (47-74)	99 (35.5%)	54 (54.6%)	3374	300K
VITAL	285	150/135	69.9 (51-83)	117 (41.1%)	70 (59.8%)	1847	300K
WHI 1	304	0/304	71.0 (52-86)	103 (33.9%)	77 (74.8%)	1868	550K, 550K duo, and 610K
WHI 2	618	0/618	71.9 (50-91)	177 (28.6%)	132 (74.6%)	1163.5	300K
WHI WGS	488	0/488	71.4 (52-89)	143 (29.3%)	99 (69.2%)	1337.5	Whole genome sequencing

Abbreviation: N: number; CRC: colorectal cancer

[a] mean (range)

[b] number (percentage of cases)

[c] number (percentages of all-cause deaths)

* All genotyping platform, except for WHI WGS, were illumina assays

Table 2. Patients characteristics and clinical features

Characteristics	Cases		Deaths, number (percentage of cases, %)	
	N	%	all-cause, N (%)	colorectal cancer, N (%)
Age, years				
<65	1345	27.5%	414 (30.8)	321(23.9)
65-69	1059	21.6%	335 (31.6)	218 (20.6)
70-74	1251	25.6%	461 (36.9)	274 (21.9)
≥75	1241	25.3%	471 (38.0)	285 (23.0)
Sex				
male	1841	37.6%	727 (39.5)	438 (23.8)
female	3055	62.4%	954 (31.2)	660 (21.6)
Cancer site				
distal	1596	32.6%	493 (30.9)	307 (19.2)
proximal	2428	49.6%	851(35.0)	556 (22.9)
rectal	726	14.8%	255 (35.1)	181(24.9)
other	146	3.0%	82 (56.2)	54 (37.0)
Cancer stage				
in situ	41	0.9%	5 (12.2)	1 (2.4)
local	1563	34.5%	278 (17.8)	74 (4.7)
regional	2366	52.2%	745 (31.5)	461 (19.5)
distant	563	12.4%	489 (86.0)	460 (81.7)
missing	363		164	102
Smoking status				
never smoker	2127	43.7%	644 (30.3)	453 (21.3)
former smoker	2245	46.1%	823 (36.7)	522 (23.3)
current smoker	494	10.1%	204 (41.3)	116 (23.5)
missing	30		10	7
Smoked cigarettes				
yes	2739	56.3%	1027 (37.5)	638 (23.3)
no	2127	43.7%	644 (30.3)	453 (21.3)
missing	30		10	7
Smoking pack-years				
<12	612	24.4%	184 (30.1)	128 (20.9)
12-24	645	25.7%	197 (30.5)	128 (19.8)
25,44	613	24.5%	232 (37.8)	145 (23.7)
≥45	635	25.4%	287 (45.2)	156 (24.6)
missing	234		127	81
Age stopped smoking				
<35	517	23.5%	154 (29.8)	111 (21.5)
35-44	516	23.5%	165 (32.0)	115 (22.3)
45-55	596	27.1%	231 (38.8)	147 (24.7)
≥55	569	25.9%	262 (46.0)	134 (23.6)
missing	570		215	321

Table 3. Hazard ratio of the association between SNPs involved in telomere maintenance and survival

Outcome	SNP	Gene	HR[a] (95% CI)	P-value	adj. P-value [b]	minor/ major allele	MAF
Overall survival	rs2075785	*TERT*	0.86 (0.76-0.98)	0.018	0.736	T/C	0.123
	rs2981096	*TERF1*	0.84 (0.70-0.99)	0.046	0.274	G/A	0.053
	rs10102030	*TNKS*	1.09 (1.01-1.18)	0.036	>0.99	T/A	0.230
	rs4939134	*TNKS1BP1*	1.08 (1.02-1.17)	0.017	>0.99	G/C	0.471
	rs1760894	*TEP1*	1.15 (1.04- 1.26)	0.004	0.224	C/T	0.222
CRC-specific survival	rs2075785	*TERT*	0.85 (0.73-0.99)	0.043	>0.99	T/C	0.123
	rs2981096	*TERF1*	0.79 (0.63-0.99)	0.040	0.2418	G/A	0.053
	rs10088969	*TNKS*	1.13 (1.02-1.25)	0.018	>0.99	C/A	0.226
	rs2229101	*TEP1*	0.78 (0.64-0.95)	0.013	0.6783	C/A	0.064
	rs251796	*TERF2*	1.13 (1.02-1.25)	0.015	0.2156	G/A	0.301

Abbreviation: CRC: colorectal cancer; HR: hazard ratio; CIs: confidence intervals; MAF: minor allele frequency
[a] adjustment for age at diagnosis, sex, study center and the first three principal components (pc)
[b] p-value is adjusted using Bonferroni method

Table 4. Hazard ratio of the association between SNPs involved in telomere maintenance and survival--- stratified by cigarette smoking

Outcome	SNP	Gene	HR [a] (95% CI)		$P_{.lrtest}$	adj. P-value [b]	minor/major allele
			Never smoker	Smoker			
Overall survival	rs56963355	*TERT*	1.27 (0.98-1.64)	0.89 (0.69-1.10)	0.045	>0.99	T/C
	rs2975842	*TERF1*	0.86 (0.77-0.96)	1.04 (0.95-1.13)	0.01	0.059	T/G
	rs35656875	*TNKS*	0.69 (0.49-0.98)	1.21 (0.87-1.54)	0.015	>0.99	A/G
	rs2409652	*PINX1*	0.90 (0.79-1.03)	1.08 (0.97-1.18)	0.03	0.867	C/T
CRC-specific survival	rs35656875	*TNKS*	0.63 (0.41-0.98)	1.40 (0.93-1.86)	0.005	>0.99	G/C
	rs56106543	*TERF2*	0.74 (0.51-1.09)	1.40 (1.04-1.76)	0.007	>0.99	C/T
	rs13259648	*PINX1*	1.27 (1.07-1.50)	0.96 (0.81-1.10)	0.014	0.174	G/A

Abbreviation: CRC: colorectal cancer; HR: hazard ratio; CIs: confidence intervals; $P_{.lrtest}$: P-value of likelihood ratio test
[a] adjustment for age at diagnosis, sex, study center and the first three principal components (pc)
[b] p-value is adjusted using Bonferroni method

Table 5. Hazard ratio of the association between SNPs involved in telomere maintenance and survival—stratified by categorical pack-years

Outcome	SNP	Gene	HR[a] (95% CI), smoking pack-years					$P_{.lrtest}$
			0	<12	12-24	25-44	≥45	
Overall survival	rs556947195	*TERT*	1.00 (0.75-1.34)	1.48 (0.87-2.08)	0.82 (0.39-1.25)	0.76 (0.33-1.19)	1.90 (1.47-2.33)	0.013
	rs2975842	*TERF1*	0.86 (0.77-0.96)	1.13 (0.89-1.36)	1.05 (0.83-1.27)	1.16 (0.94-1.38)	0.98 (0.76-1.20)	0.033
	rs251797	*TERF2*	0.99 (0.88-1.12)	1.08 (0.84-1.33)	1.06 (0.82-1.30)	1.25 (1.01-1.49)	0.78 (0.54-1.02)	0.019
	rs67456872	*TNKS*	0.68 (0.48-0.97)	2.42 (1.00-3.84)	0.58 (0.17-0.99)	1.59 (1.18-2.00)	0.96 (0.55-1.37)	0.002
	rs76990680	*TNKS*	1.15 (1.02-1.30)	0.82 (0.63-1.02)	1.24 (0.99-1.49)	0.97 (0.72-1.22)	0.95 (0.70-1.20)	0.096
CRC-specific survival	rs2242652	*TERT*	0.89 (0.73-1.07)	0.76 (0.48-1.03)	1.40 (0.93-1.87)	0.82 (0.35-1.29)	1.29 (0.82-1.76)	0.025
	rs2853690	*TERT*	1.12 (0.93-1.33)	1.17 (0.81-1.53)	0.65 (0.41-0.89)	1.30 (1.06-1.54)	1.05 (0.81-1.29)	0.042
	rs153045	*TERF2*	1.10 (0.95-1.28)	1.16 (0.83-1.48)	1.18 (0.85-1.51)	0.74 (0.41-1.07)	1.41 (1.08-1.74)	0.024
	rs10091836	*PINX1*	0.88 (0.77-1.00)	0.78 (0.59-0.98)	1.19 (0.90-1.48)	1.00 (0.71-1.29)	1.26 (0.97-1.55)	0.011
	rs67456872	*TNKS*	0.63 (0.40-0.97)	2.67 (0.91-4.43)	0.81 (0.16-1.46)	2.16 (1.51-2.81)	0.95 (0.30-1.60)	0.002
	rs7200950	***ACD***	**1.11 (0.83-1.48)**	**1.65 (0.91-2.39)**	**0.69 (0.24-1.14)**	**1.45 (1.00-1.90)**	**0.52 (0.07-0.97)**	**0.016***

Abbreviation: CRC: colorectal cancer; HR: hazard ratio; CI: confidence intervals; $P_{.lrtest}$: P-value of likelihood ratio test

[a] adjustment for age at diagnosis, sex, study center and the first three principal components (pc)

* adjusted p-value is 0.0495; other adjusted p-values are not significant (>0.05)

Table 6. Hazard ratio of the association between SNPs involved in telomere maintenance and survival — stratified by smoking pack-years (≥45 vs. 0)

Outcome	SNP	Gene	HR[a] (95% CI), smoking pack-years		$P_{.lrtest}$	adj. P-value [b]
			0	≥45		
Overall survival	rs556947195	TERT	1.01 (0.76-1.34)	1.94 (1.20-2.68)	0.0084	0.336
	rs251797	TERF2	0.99 (0.88-1.12)	0.77 (0.62-0.92)	0.0323	0.452
	rs1865493	**TERF2IP**	**0.98 (0.81-1.19)**	**0.57 (0.36-0.77)**	**0.0056**	**0.045***
	rs2409655	PINX1	0.92 (0.81-1.04)	1.17 (0.96-1.38)	0.0279	0.809
	rs17677264	TNKS	1.19 (0.95-1.51)	0.64 (0.38-0.91)	0.0078	0.624
	rs79784010	TNKS2	1.02 (0.80-1.29)	0.61 (0.34-0.89)	0.0434	0.434
CRC-specific survival	**rs73202875**	**TNKS**	**1.37 (1.07-1.75)**	**0.46 (0.19-0.74)**	**3.00×10^{-4}**	**0.024***

Abbreviation: CRC: colorectal cancer; HR: hazard ratio; CI: confidence intervals; $P_{.lrtest}$: P-value of likelihood ratio test

[a] adjustment for age at diagnosis, sex, study center and the first three principal components (pc)

[b] p-value is adjusted using Bonferroni method

Table 7. Hazard ratio of the association between SNPs involved in telomere maintenance and survival--- stratified by sex

Outcome	SNP	Gene	HR [a] (95% CIs)		$P._{lrtest}$	adj. P-value [b]
			Female	Male		
Overall survival	**rs75676021**	***POT1***	**1.21 (1.01-1.45)**	**0.77 (0.60-0.95)**	**0.002**	**0.023***
	rs2853685	*TERT*	1.14 (1.03-1.26)	0.89 (0.79-1.00)	0.002	0.096
	rs2975843	***TERF1***	**1.08 (0.99-1.18)**	**0.84 (0.75-0.92)**	**$3.00x10^{-4}$**	**0.002***
	rs73615082	*TERF2IP*	0.76 (0.62-0.93)	1.08 (0.85-1.30)	0.019	0.154
	rs4840518	*PINX1*	1.21 (1.02-1.43)	0.90 (0.72-1.08)	0.026	0.748
	rs77103162	*TNKS1BP1*	0.93 (0.80-1.08)	1.17 (0.99-1.34)	0.0337	0.506
	rs35259162	*TEP1*	0.83 (0.73-0.94)	1.02 (0.88-1.16)	0.0293	>0.99
	rs3950296	*TERC*	0.91 (0.82-1.02)	1.09 (0.96-1.22)	0.0278	0.139
CRC-specific survival	**rs74429678**	***POT1***	**1.33 (1.07-1.65)**	**0.75 (0.52-0.97)**	**0.0019**	**0.019***
	rs2736115	*TERT*	1.21 (1.07-1.37)	0.90 (0.76-1.04)	0.0027	0.108
	rs2975843	***TERF1***	**1.12 (1.00-1.25)**	**0.83 (0.71-0.94)**	**$6.00x10^{-4}$**	**0.004***
	rs73615082	*TERF2IP*	0.77 (0.60-0.98)	1.12 (0.83-1.42)	0.0391	0.313
	rs10503412	*PINX1*	1.21 (1.00-1.46)	0.78 (0.58-0.97)	0.0051	0.148
	rs4416825	*TNKS*	0.97 (0.83-1.14)	1.24 (1.01-1.47)	0.0482	>0.99
	rs1760895	*TEP1*	1.19 (1.02-1.40)	0.86 (0.67-1.06)	0.0195	0.995
	rs9876206	*TERC*	0.89 (0.78-1.01)	1.12 (0.95-1.28)	0.0252	0.126

Abbreviation: CRC: colorectal cancer; HR: hazard ratio; CI: confidence intervals; $P._{lrtest}$: P-value of likelihood ratio test

[a] adjustment for age at diagnosis, study center and the first three principal components (pc)

[b] p-value is adjusted using Bonferroni method

References

1. Key Statistics for Colorectal Cancer [Internet]. [cited 2018 Mar 5]. Available from: https://www.cancer.org/cancer/colon-rectal-cancer/about/key-statistics.html

2. Walther A, Johnstone E, Swanton C, Midgley R, Tomlinson I, Kerr D. Genetic prognostic and predictive markers in colorectal cancer. Nat Rev Cancer. 2009;9:489–99.

3. Garcia CK, Wright WE, Shay JW. Human diseases of telomerase dysfunction: insights into tissue aging. Nucleic Acids Res. 2007;35:7406–16.

4. Wentzensen IM, Mirabello L, Pfeiffer RM, Savage SA. The Association of Telomere Length and Cancer: A Meta-Analysis. Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol. 2011;20:1238–50.

5. Blackburn EH, Epel ES, Lin J. Human telomere biology: A contributory and interactive factor in aging, disease risks, and protection. Science. 2015;350:1193–8.

6. Lange T de. Shelterin: the protein complex that shapes and safeguards human telomeres. Genes Dev. 2005;19:2100–10.

7. Collins K, Mitchell JR. Telomerase in the human organism. Oncogene. 2002;21:564–79.

8. Savage SA, Stewart BJ, Eckert A, Kiley M, Liao JS, Chanock SJ. Genetic variation, nucleotide diversity, and linkage disequilibrium in seven telomere stability genes suggest that these genes may be under constraint. Hum Mutat. 2005;26:343–50.

9. Latifovic L, Peacock SD, Massey TE, King WD. The Influence of Alcohol Consumption, Cigarette Smoking, and Physical Activity on Leukocyte Telomere Length. Cancer Epidemiol Prev Biomark. 2016;25:374–80.

10. Astuti Y, Wardhana A, Watkins J, Wulaningsih W. Cigarette smoking and telomere length: A systematic review of 84 studies and meta-analysis. Environ Res. 2017;158:480–9.

11. Boardman LA, Johnson RA, Viker KB, Hafner KA, Jenkins RB, Riegert-Johnson DL, et al. Correlation of Chromosomal Instability, Telomere Length and Telomere Maintenance in Microsatellite Stable Rectal Cancer: A Molecular Subclass of Rectal Cancer. PLoS ONE [Internet]. 2013 [cited 2018 Apr 6];8. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3836975/

12. Roger L, Jones RE, Heppel NH, Williams GT, Sampson JR, Baird DM. Extensive Telomere Erosion in the Initiation of Colorectal Adenomas and Its Association With Chromosomal Instability. JNCI J Natl Cancer Inst. 2013;105:1202–11.

13. Mocellin S, Verdi D, Pooley KA, Landi MT, Egan KM, Baird DM, et al. Telomerase reverse transcriptase locus polymorphisms and cancer risk: a field synopsis and meta-analysis. J Natl Cancer Inst. 2012;104:840–54.

14. Peters U, Hutter CM, Hsu L, Schumacher FR, Conti DV, Carlson CS, et al. Meta-analysis of new genome-wide association studies of colorectal cancer risk. Hum Genet. 2012;131:217–34.

15. 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. J Natl Cancer Inst. 2015;107.

16. 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.

17. Karami S, Han Y, Pande M, Cheng I, Rudd J, Pierce BL, et al. Telomere structure and maintenance gene variants and risk of five cancer types. Int J Cancer. 2016;139:2655–70.

18. Tatsumoto N, Hiyama E, Murakami Y, Imamura Y, Shay JW, Matsuura Y, et al. High Telomerase Activity Is an Independent Prognostic Indicator of Poor Outcome in Colorectal Cancer. Clin Cancer Res. 2000;6:2696–701.

19. Harris HR, Vivo ID, Titus LJ, Vitonis AF, Wong JYY, Cramer DW, et al. Genetic variation in telomere maintenance genes in relation to ovarian cancer survival. Int J Mol Epidemiol Genet. 2012;3:252–61.

20. Jung Seok Won, Park Neung Hwa, Shin Jung Woo, Park Bo Ryung, Kim Chang Jae, Lee Jong-Eun, et al. Prognostic impact of telomere maintenance gene polymorphisms on hepatocellular carcinoma patients with chronic hepatitis B. Hepatology. 2014;59:1912–20.

21. Lin X, Gu J, Lu C, Spitz MR, Wu X. Expression of Telomere-Associated Genes as Prognostic Markers for Overall Survival in Patients with Non–Small Cell Lung Cancer. Clin Cancer Res. 2006;12:5720–5.

22. Yuan P, Cao J, Abuduwufuer A, Wang L-M, Yuan X-S, Lv W, et al. Clinical Characteristics and Prognostic Significance of TERT Promoter Mutations in Cancer: A Cohort Study and a Meta-Analysis. PLOS ONE. 2016;11:e0146803.

23. Mosrati MA, Malmström A, Lysiak M, Krysztofiak A, Hallbeck M, Milos P, et al. TERT promoter mutations and polymorphisms as prognostic factors in primary glioblastoma. Oncotarget. 2015;6:16663–73.

24. Röth A, Harley CB, Baerlocher GM. Imetelstat (GRN163L)--telomerase-based cancer therapy. Recent Results Cancer Res Fortschritte Krebsforsch Progres Dans Rech Sur Cancer. 2010;184:221–34.

25. Bernardes de Jesus B, Vera E, Schneeberger K, Tejera AM, Ayuso E, Bosch F, et al. Telomerase gene therapy in adult and old mice delays aging and increases longevity without increasing cancer. EMBO Mol Med. 2012;4:691–704.

26. 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.

27. Gao X, Starmer J, Martin ER. A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms. Genet Epidemiol. 2008;32:361–9.

28. Jannuzzi AT, Karaman E, Oztas E, Yanar HT, Özhan G. Telomerase Reverse Transcriptase (TERT) Gene Variations and Susceptibility of Colorectal Cancer. Genet Test Mol Biomark. 2015;19:692–7.

29. Rachakonda PS, Hosen I, de Verdier PJ, Fallah M, Heidenreich B, Ryk C, et al. TERT promoter mutations in bladder cancer affect patient survival and disease recurrence through modification by a common polymorphism. Proc Natl Acad Sci U S A. 2013;110:17426–31.

30. Catarino R, Araújo A, Coelho A, Gomes M, Nogueira A, Lopes C, et al. Prognostic significance of telomerase polymorphism in non-small cell lung cancer. Clin Cancer Res Off J Am Assoc Cancer Res. 2010;16:3706–12.

31. Shen J, Gammon MD, Terry MB, Bradshaw PT, Wang Q, Teitelbaum SL, et al. Genetic polymorphisms in telomere pathway genes, telomere length, and breast cancer survival. Breast Cancer Res Treat. 2012;134:393–400.

32. Sun Y, Tao W, Huang M, Wu X, Gu J. Genetic variants in telomere-maintenance genes are associated with ovarian cancer risk and outcome. J Cell Mol Med. 2017;21:510–8.

33. Phipps AI, Passarelli MN, Chan AT, Harrison TA, Jeon J, Hutter CM, et al. Common genetic variation and survival after colorectal cancer diagnosis: a genome-wide analysis. Carcinogenesis. 2016;37:87–95.

34. Hackett JA, Greider CW. Balancing instability: dual roles for telomerase and telomere dysfunction in tumorigenesis. Oncogene. 2002;21:619–26.

35. Gertler R, Rosenberg R, Stricker D, Friederichs J, Hoos A, Werner M, et al. Telomere Length and Human Telomerase Reverse Transcriptase Expression As Markers for Progression and Prognosis of Colorectal Carcinoma. J Clin Oncol. 2004;22:1807–14.

36. Jia H, Wang Z. Telomere Length as a Prognostic Factor for Overall Survival in Colorectal Cancer Patients. Cell Physiol Biochem. 2016;38:122–8.

37. Chen Y, Qu F, He X, Bao G, Liu X, Wan S, et al. Short leukocyte telomere length predicts poor prognosis and indicates altered immune functions in colorectal cancer patients. Ann Oncol. 2014;25:869–76.

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

39. Bertorelle R, Briarava M, Rampazzo E, Biasini L, Agostini M, Maretto I, et al. Telomerase is an independent prognostic marker of overall survival in patients with colorectal cancer. Br J Cancer. 2013;108:278–84.

40. Kojima Kensaku, Hiyama Eiso, Otani Keiko, Ohtaki Megu, Fukuba Ikuko, Fukuda Emi, et al. Telomerase activation without shortening of telomeric 3′-overhang is a poor prognostic factor in human colorectal cancer. Cancer Sci. 2010;102:330–5.

41. Fernández-Marcelo T, Sánchez-Pernaute A, Pascua I, Juan CD, Head J, Torres-García A-J, et al. Clinical Relevance of Telomere Status and Telomerase Activity in Colorectal Cancer. PLOS ONE. 2016;11:e0149626.

42. Liu H, Liu Q, Ge Y, Zhao Q, Zheng X, Zhao Y. hTERT promotes cell adhesion and migration independent of telomerase activity. Sci Rep [Internet]. 2016 [cited 2018 May 1];6. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4789728/

43. Ghosh A, Saginc G, Leow SC, Khattar E, Shin EM, Yan TD, et al. Telomerase directly regulates NF-κB-dependent transcription. Nat Cell Biol. 2012;14:1270–81.

44. Li S, Crothers J, Haqq CM, Blackburn EH. Cellular and gene expression responses involved in the rapid growth inhibition of human cancer cells by RNA interference-mediated depletion of telomerase RNA. J Biol Chem. 2005;280:23709–17.

45. Naito Y, Takagi T, Handa O, Ishikawa T, Matsumoto N, Yoshida N, et al. Telomerase activity and expression of telomerase RNA component and catalytic subunits in precancerous and cancerous colorectal lesions. Tumour Biol J Int Soc Oncodevelopmental Biol Med. 2001;22:374–82.

46. Kawanishi S, Oikawa S. Mechanism of telomere shortening by oxidative stress. Ann N Y Acad Sci. 2004;1019:278–84.

47. Ambatipudi S, Cuenin C, Hernandez-Vargas H, Ghantous A, Le Calvez-Kelm F, Kaaks R, et al. Tobacco smoking-associated genome-wide DNA methylation changes in the EPIC study. Epigenomics. 2016;8:599–618.

48. Lee KWK, Pausova Z. Cigarette smoking and DNA methylation. Front Genet [Internet]. 2013 [cited 2018 Apr 23];4. Available from: https://www.frontiersin.org/articles/10.3389/fgene.2013.00132/full

49. Pellatt AJ, Wolff RK, Lundgreen A, Cawthon R, Slattery ML. Genetic and lifestyle influence on telomere length and subsequent risk of colon cancer in a case control study. Int J Mol Epidemiol Genet. 2012;3:184–94.

50. Patel CJ, Manrai AK, Corona E, Kohane IS. Systematic correlation of environmental exposure and physiological and self-reported behaviour factors with leukocyte telomere length. Int J Epidemiol. 2017;46:44–56.

51. Sung YJ, Winkler TW, de las Fuentes L, Bentley AR, Brown MR, Kraja AT, et al. A Large-Scale Multi-ancestry Genome-wide Study Accounting for Smoking Behavior Identifies Multiple Significant Loci for Blood Pressure. Am J Hum Genet. 2018;102:375–400.

52. Lange T de. How Shelterin Solves the Telomere End-Protection Problem. Cold Spring Harb Symp Quant Biol. 2010;75:167–77.

53. Savage Sharon A., Stewart Brian J., Eckert Andrew, Kiley Maureen, Liao Jason S., Chanock Stephen J. Genetic variation, nucleotide diversity, and linkage disequilibrium in seven telomere stability genes suggest that these genes may be under constraint. Hum Mutat. 2005;26:343–50.

54. 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 [Internet]. 2014 [cited 2018 May 10];9. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3912164/

55. Pan W, Du J, Shi M, Jin G, Yang M. Short leukocyte telomere length, alone and in combination with smoking, contributes to increased risk of gastric cancer or esophageal squamous cell carcinoma. Carcinogenesis. 2017;38:12–8.

56. Getliffe K. M., Al Dulaimi D., Martin-Ruiz C., Holder R. L., Von Zglinicki T., Morris A., et al. Lymphocyte telomere dynamics and telomerase activity in inflammatory bowel disease: effect of drugs and smoking. Aliment Pharmacol Ther. 2005;21:121–31.

57. Xinarianos G, Scott FM, Liloglou T, Prime W, Callaghan J, Gosney JR, et al. Telomerase activity in non-small cell lung carcinomas correlates with smoking status. Int J Oncol. 1999;15:961–6.

58. Targowski T, Jahnz-Rózyk K, Szkoda T, From S, Rozyńska R, Płusa T. [Influence of nicotine addiction on telomerase activity in malignant non-small cell lung tumors]. Przegl Lek. 2005;62:1043–6.

59. Gong J, Hutter CM, Newcomb PA, Ulrich CM, Bien SA, Campbell PT, et al. Genome-Wide Interaction Analyses between Genetic Variants and Alcohol Consumption and Smoking for Risk of Colorectal Cancer. PLoS Genet. 2016;12:e1006296.

60. Gao Y, Hayes RB, Huang W-Y, Caporaso NE, Burdette L, Yeager M, et al. DNA repair gene polymorphisms and tobacco smoking in the risk for colorectal adenomas. Carcinogenesis. 2011;32:882–7.

61. Barry EL, Poole EM, Baron JA, Makar KW, Mott LA, Sandler RS, et al. CYP2C9 variants increase risk of colorectal adenoma recurrence and modify associations with smoking but not aspirin treatment. Cancer Causes Control CCC. 2013;24:47–54.

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

63. Zhu H, Wang X, Gutin B, Davis CL, Keeton D, Thomas J, et al. Leukocyte Telomere Length in Healthy Caucasian and African-American Adolescents: Relationships with Race, Sex, Adiposity, Adipokines, and Physical Activity. J Pediatr. 2011;158:215–20.

64. Gardner M, Bann D, Wiley L, Cooper R, Hardy R, Nitsch D, et al. Gender and telomere length: Systematic review and meta-analysis. Exp Gerontol. 2014;51:15–27.

Appendix: Supplementary Material

1. Supplementary methods

1.1 Details on included study

Diet, Activity, and Lifestyle Study (DALS) (1-3)

DALS (Diet, Activity, and Lifestyle Study) was a population-based, case-control study of colon cancer. Participants were recruited between 1991 and 1994 from 3 locations: the Kaiser Permanente Medical Care Program of Northern California, an 8-county area in Utah, and the metropolitan Twin Cities area of Minnesota. Eligibility criteria for cases included age at diagnosis between 30 and 79 years; diagnosis with first primary colon cancer (International Classification of Diseases for Oncology second edition codes 18.0 and 18.2–18.9) between October 1, 1991, and September 30, 1994; English speaking; and competency to complete the interview. Individuals with cancer of the rectosigmoid junction or rectum were excluded, as were those with a pathology report noting familial adenomatous polyposis, Crohn's disease, or ulcerative colitis. A rapid-reporting system was used to identify all incident cases of colon cancer, resulting in the majority of cases being interviewed within 4 months of diagnosis. Controls from the Kaiser Permanente Medical Care Program were selected randomly from membership lists. In Utah, controls younger than 65 years of age were selected randomly through random-digit dialing and driver's license lists. Controls, 65 years of age and older, were selected randomly from Health Care Financing Administration lists. In Minnesota, controls were identified from Minnesota driver's licenses or state identification lists. Controls were matched to cases by 5-year age groups and sex. The set I scan consisted of a subset of the study designed earlier, from Utah, Minnesota, and the Kaiser Permanente Medical Care Program, and was

restricted to subjects who self-reported as white non-Hispanic. The set 2 scan consisted of subjects from Utah and Minnesota who were not genotyped in set 1. Set 2 was restricted to subjects who self-reported as white non-Hispanic and those who had appropriate consent to post data to dbGaP.

Health Professionals Follow-up Study (HPFS)

The HPFS (Health Professionals Follow-up Study) is a parallel prospective study to the NHS (Nurses' Health Study) (4). The HPFS cohort comprised 51,529 men who, in 1986, responded to a mailed questionnaire. The participants were US male dentists, optometrists, osteopaths, podiatrists, pharmacists, and veterinarians born between 1910 and 1946. Participants provided information on health-related exposures, including current and past smoking history, age, weight, height, diet, physical activity, aspirin use, and family history of colorectal cancer. Colorectal cancer and other outcomes were reported by participants or next-of-kin and were followed up through review of the medical and pathology record by physicians. Overall, more than 97% of self-reported colorectal cancers were confirmed by medical record review. Information was abstracted on histology and primary location. Incident cases were defined as those occurring after the subject provided the blood sample. Prevalent cases were defined as those occurring after enrollment in the study but before the subject provided the blood sample. Follow-up evaluation has been excellent, with 94% of the men responding to date. Colorectal cancer cases were ascertained through January 1, 2008. In 1993–1995, 18,825 men in the HPFS mailed blood samples by overnight courier, which were aliquoted into buffy coat and stored in liquid nitrogen. Between 2001 and 2004, 13,956 men in the HPFS who had not provided a blood sample previously mailed in a swish-and-spit sample of buccal cells. Incident cases were defined as those occurring after the subject provided a blood or buccal sample. Prevalent cases were defined

as those occurring after enrollment in the study in 1986, but before the subject provided either a blood or buccal sample; these prevalent cases were excluded from our analyses. After excluding participants with histories of cancer (except non-melanoma skin cancer), ulcerative colitis, or familial polyposis, 2 case-control sets were constructed from which DNA was isolated from either buffy coat or buccal cells for genotyping, as follows: 1) a case-control set with cases of colorectal cancer matched to randomly selected controls who provided a blood sample and were free of colorectal cancer at the same time the colorectal cancer was diagnosed in the cases; and 2) a case-control set with cases of colorectal cancer matched to randomly selected controls who provided a buccal sample and were free of colorectal cancer at the same time the colorectal cancer was diagnosed in the case. For both case-control sets, matching criteria included year of birth (within 1 year) and month/year of blood or buccal cell sampling (within 6 months). Cases were pair-matched 1:1, 1:2, or 1:3 with control participant(s).

Nurses' Health Study (NHS)

The NHS cohort began in 1976 when 121,700 married female registered nurses age 30–55 years returned the initial questionnaire that ascertained a variety of important health-related exposures (5). Since 1976, follow-up questionnaires have been mailed every 2 years. Colorectal cancer and other outcomes were reported by participants or next-of-kin and followed up through review of the medical and pathology record by physicians. Overall, more than 97% of self-reported colorectal cancers were confirmed by medical-record review. Information was abstracted on histology and primary location. The rate of follow-up evaluation has been high: as a proportion of the total possible follow-up time, follow-up evaluation has been more than 92%. Colorectal

cancer cases were ascertained through June 1, 2008. In 1989–1990, 32,826 women in NHS I mailed blood samples by overnight courier, which were aliquoted into buffy coat and stored in liquid nitrogen. In 2001–2004, 29,684 women in NHS I who did not previously provide a blood sample mailed a swish-and-spit sample of buccal cells. Incident cases were defined as those occurring after the subject provided a blood or buccal sample. Prevalent cases were defined as those occurring after enrollment in the study in 1976 but before the subject provided either a blood or buccal sample; these prevalent cases were excluded from our analyses. After excluding participants with histories of cancer (except non-melanoma skin cancer), ulcerative colitis, or familial polyposis, 2 case-control sets were constructed from which DNA was isolated from either buffy coat or buccal cells for genotyping: 1) a case-control set with cases of colorectal cancer matched to randomly selected controls who provided a blood sample and were free of colorectal cancer at the same time the colorectal cancer was diagnosed in the case; and 2) a case-control set with cases of colorectal cancer matched to randomly selected controls who provided a buccal sample and were free of colorectal cancer at the same time the colorectal cancer was diagnosed in the cases. For both case-control sets, matching criteria included year of birth (within 1 year) and month/year of blood or buccal cell sampling (within 6 months). Cases were pair matched 1:1, 1:2, or 1:3 with control participant(s).

Physicians' Health Study (PHS)

The PHS (Physicians' Health Study) was established as a randomized, double-blind, placebo-controlled trial of aspirin and β-carotene among 22,071 healthy US male physicians, between 40 and 84 years of age, in 1982 (6)(7). Participants completed 2 mailed questionnaires before being assigned randomly, additional questionnaires at 6 and 12 months, and questionnaires annually

thereafter. In addition, participants were sent postcards at 6 months to ascertain status. From August 1982 to December 1984, there were 14,916 baseline blood samples collected from the physicians during the run-in phase before randomization. When participants reported a diagnosis of cancer, medical records and pathology reports were reviewed by study physicians who were blinded to exposure data. Among those who provided baseline blood samples, colorectal cases were ascertained through March 31, 2008, and controls were matched on age (within 1 year for younger participants, up to 5 years for older participants) and smoking status (never, past, current). Cases were pair-matched 1:1, 1:2, or 1:3 with control participant(s). Because of DNA availability, samples were genotyped in 2 batches on the same platform at the same genotyping center at different time points.

Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial (PLCO)

The PLCO (Prostate, Lung, Colorectal Cancer, and Ovarian Cancer Screening Trial) enrolled 154,934 participants (men and women, aged between 55 and 74 y) at 10 centers into a large, randomized, 2-arm trial to determine the effectiveness of screening to reduce cancer mortality. Sequential blood samples were collected from participants assigned to the screening arm. Participation was 93% at the baseline blood draw. In the observational (control) arm, buccal cells were collected via mail using the swish-and-spit protocol; the participation rate was 65%. Details of this study have been described previously (8-9) and are available online (http://dcp.cancer.gov/plco).

The set 1 scan included a subset of 577 colon cancer cases self-reported as being non-Hispanic white with available DNA samples, questionnaire data, and appropriate consent for ancillary

epidemiologic studies. Cases were excluded if they had a history of inflammatory bowel disease, polyps, polyposis syndrome, or cancer (excluding basal or squamous cell skin cancer). Controls originated from the Cancer Genetic Markers of Susceptibility prostate cancer scan (10-11)(all male) and the GWAS of Lung Cancer and Smoking (12) (enriched for smokers), along with an additional 92 non-Hispanic white female controls. Set 1 samples have been excluded from this analysis. For the set 2 scan, cases were individuals with colorectal cancer from both arms of the trial who were not already included in set 1. Samples were excluded if participants did not sign appropriate consent forms, if DNA was unavailable, if baseline questionnaire data with follow-up evaluation were unavailable, if they had a history of colon cancer before the trial, if they had a rare cancer, if they were already in a colon GWAS, or if they were a control in the prostate or lung populations. Controls were frequency-matched 1:1 to cases without replacement, and cases were not eligible to be controls. Matching criteria were age at enrollment (2-year blocks), enrollment date (2-year blocks), sex, race/ethnicity, trial arm, and study year of diagnosis (ie, controls must be cancer free into the case's year of diagnosis).

Postmenopausal Hormones Supplementary Study to the CCFR (PMH-CCFR)

Eligible case patients included all female residents, ages 50–74 years, residing in the 13 counties in Washington State, reporting to the Cancer Surveillance, Epidemiology and End Results program, who were newly diagnosed with invasive colorectal adenocarcinoma (ICD-O C18.0, C18.2–C18.9, C19.9, C20.0–C20.9) between October 1998 and February 2002 (13). Eligibility for all individuals was limited to those who were English speaking with available telephone numbers, through which they could be contacted. On average, cases were identified within 4 months of diagnosis. The overall response proportion of eligible cases identified was 73%.

Community-based controls were selected randomly according to age distribution (in 5-year age intervals) of the eligible cases by using lists of licensed drivers from the Washington State Department of Licensing for individuals, ages 50–64 years, and rosters from the Health Care Financing Administration (now the Centers for Medicare and Medicaid), for individuals older than age 64. The overall response proportion of eligible controls was 66%. In GECCO, samples with sufficient DNA extracted from blood were genotyped. Only participants who were not part of the CCFR Seattle site were included in the sample set.

VITamins And Lifestyle (VITAL)

The VITAL (VITamins And Lifestyle) cohort comprised 77,721 Washington State men and women aged 50–76 years, recruited from 2000 to 2002, to investigate the association of supplement use and lifestyle factors with cancer risk. Subjects were recruited by mail, from October 2000 to December 2002, using names purchased from a commercial mailing list. All subjects completed a 24-page questionnaire and buccal-cell specimens for DNA were self-collected by 70% of the participants. Subjects were followed for cancer by linkage to the western Washington Surveillance, Epidemiology and End Results (SEER) cancer registry and were censored when they moved out of the area covered by the registry or at time of death. Details of this study have been described previously (14). In GECCO, a nested case-control set was genotyped. Samples included colorectal cancer cases with DNA, excluding subjects with colorectal cancer before baseline, in situ cases, (large cell) neuroendocrine carcinoma, squamous cell carcinoma, carcinoid tumor, Goblet-cell carcinoid, and any type of lymphoma, including non-Hodgkin, Mantle cell, large B-cell, or follicular lymphoma. Controls were matched on age at enrollment (within 1 year), enrollment date (within 1 year), sex, and race/ethnicity. One control

was selected randomly per case among all controls who matched according to the 4 factors described earlier and for whom the control follow-up time was greater than the follow-up time of the case until diagnosis.

Women's Health Initiative (WHI)

The WHI (Women's Health Initiative) is a long-term health study of 161,808 post-menopausal women aged 50–79 years at 40 clinical centers throughout the United States. WHI comprised a clinical trial arm, an observational study (OS) arm, and several extension studies. The details of WHI have been described previously (15,16) and are available online (https://cleo.whi.org/SitePages/Home.aspx). In GECCO, set 1 cases were selected from the September 12, 2005, database and comprised centrally adjudicated colon cancer cases from the OS arm who self-reported as white. Controls were first selected among controls previously genotyped as part of a hip fracture GWAS conducted within the WHI OS arm and matched to cases on age (within 3 years), enrollment date (within 365 days), hysterectomy status, and prevalent conditions at baseline. For 37 cases, there was no control match in the hip fracture GWAS. For these participants, we identified a matched control in the WHI OS arm based on the same criteria. In the set 2 scan, cases were selected from the August 2009 database and comprised centrally adjudicated colon and colorectal cancer cases from the OS and clinical trial arms who were not genotyped in set 1. In addition, case and control participants were subject to the following exclusion criteria: a prior history of colorectal cancer at baseline, institutional review board approval not available for data submission into dbGaP, and insufficient DNA available. Matching criteria included age (within 3 years), race/ethnicity, WHI date (within 3 years), WHI Calcium and Vitamin D study date (within 3 years), and randomization arms (OS

flag, hormone therapy assignments, dietary modification assignments, calcium/vitamin D assignments). In addition, study participants were matched on randomization centers, with analytic adjustment made for the 4 regions of randomization centers. Each case was matched with 1 control (1:1) who met the matching criteria exactly (i.e., cases and controls were matched on trial assignments and on randomization within these trials; in the rare event that a case is in 2 trials, she was matched to a control in the same two trials). Control selection was performed in a time-forward manner, selecting one control for each case first from the risk set at the time of the case's event. The matching algorithm was allowed to select the closest match based on a criterion to minimize an overall distance measure (17). Each matching factor was given the same weight. Additional available controls who were genotyped as part of the hip fracture GWAS were included to improve power.

1.2 Details on Genotyping Platform and Quality Assurance and Quality Control (QA/QC)

All analyses were based on genotyped data (except for WHI_WGS) generated from genome-wide association scans and imputation to HapMap II. We note that genotyping for some cohorts was conducted at two different time points (i.e., sets 1 and 2) based on the availability of funds and samples. We accounted for each set in the statistical analysis. Phase one genotyping of DALS Set 1 and WHI Set 1 was done using Illumina HumanHap 550K/610K and Illumina 550Kduo/610K, respectively, and has been described previously (18). DALS Set 2, PMH-CCFR, PLCO Set 2, VITAL, and WHI Set 2 were genotyped using Illumina HumanCytoSNP. HPFS, NHS, and DACHS Set 2 were genotyped using Illumina HumanOmniExpress. WHI WGS set was sequenced by whole genome sequencing.

DNA was extracted from blood samples or, for a subset of HPFS, NHS, and PLCO samples, and for all VITAL samples, from buccal cells, using conventional methods. All studies included 1 to 6% blinded duplicates to monitor quality of the genotyping. All individual-level genotype data were managed, and underwent quality assurance and quality control (QA/QC) at University of Washington Genetics Coordinating Center (HPFS and NHS) or the GECCO Coordinating Center at the Fred Hutchinson Cancer Research Center (all other studies). Samples were excluded based on call rate (≤97%), heterozygosity, unexpected duplicates and gender discrepancy. All analyses were restricted to samples clustering with the Utah residents with Northern and Western European ancestry from the CEPH collection (CEU) population in principal component analysis, (19) including the HapMap II populations as reference. Single nucleotide polymorphisms (SNPs) were excluded if they were reported or observed as not performing consistently across platforms. Additionally, genotyped SNPs were excluded based on call rate (< 98%), lack of Hardy-Weinberg Equilibrium in controls (HWE, $P < 1x10\text{-}4$), and minor allele frequency (MAF < 5% for WHI Set 1, DALS Set 1; MAF < 5 / # of samples for each other study). As imputation of genotypes is established as standard practice in the genetic association analysis, all autosomal SNPs of each study were imputed to the CEU population in HapMap II release 24 using MACH. (20) Imputed data were merged with genotype data such that genotype data were used if a SNP had both types of data, unless there was a difference in terms of reference allele frequency (> 0.1) or position (> 100 base pairs), in which case imputed data were used. We calculated R^2 as a measurement of imputation accuracy. After imputation and quality control (QC) analyses, a total of about 6578 SNPs were included in the study and 1871 common SNPs were used for analysis. In the statistical analyses, both genotyped and imputed SNPs were examined as continuous variables ranging from 0-2. Briefly, under the log-additive model, the statistical effect of a

homozygous variant genotype is assumed to be twice the statistical effect of a heterozygous genotype on a logit-scale. This is equivalent to considering genotype according to dosage or number of variant alleles (0, 1 and 2) and evaluating its contribution to the model as a continuous covariate. For imputed genotypes, we obtained the posterior probabilities for heterozygous and homozygous variant genotypes from the MACH imputation program to calculate the expected dosage as 2Pr(Genotype=AA) + Pr(Genotype=Aa). Because the posterior probabilities are constrained between 0 and 1, the expected dosage will be between 0 and 2. We have previously shown that the expected dosage provides a valid inference of the actual number of variant alleles. (21)

2. Supplementary tables

Table S1. Characteristics of included genes

Gene	Description	Position	SNPs (N)	M_{eff_G}	MAF, median(range)	R^2, mean
ACD	adrenocortical dysplasia homolog	16q22.1	4	3	0.099 (0.051-0.488)	0.906
PINX1	PIN2/TERF1-interacting telomerase inhibitor 1	8p23.1	378	29	0.234 (0.050-0.482)	0.984
POT1	protection of telomeres 1	7q31.33	276	10	0.402 (0.052-0.418)	0.968
TEP1	telomerase-associated protein 1	14q11.2	130	51	0.167 (0.051-0.491)	0.918
TERF1	telomeric repeat binding factor 1	8q21.11	51	6	0.308 (0.053-0.454)	0.997
TERF2	telomeric repeat binding factor 2	16q22.1	23	14	0.268 (0.061-0.398)	0.897
TERF2IP	TERF2 Interacting Protein	16q23.1	15	8	0.138 (0.050-0.306)	0.892
TERT	telomerase reverse transcriptase	5p15.33	62	40	0.267 (0.053-0.494)	0.765
TINF2	TERF1-interacting nuclear factor 2	14q12	2	2	0.144, 0.192	0.954
TNKS	tankyrase	8p23.1	715	80	0.172 (0.050-0.433)	0.973
TNKS1BP1	tankryase 1 binding protein 1	11q12.1	33	15	0.263 (0.051-0.480)	0.954
TNKS2	tankyrase 2	10q23.32	147	10	0.182 (0.060-0.488)	0.968
TERC	telomerase RNA component	3q26.2	35	5	0.259 (0.050-0.478)	0.958

Abbreviation: M_{eff_G}: effective number of SNPs used for gene-level Bonferroni correction; MAF: minor allele frequency; #: number

Table S2. Hazard ratio of the association between SNPs involved in telomere maintenance and overall survival and CRC-specific survival--- stratified by smoking status[+]

Outcome	SNP	gene	HR [a] (95% CI)			$P_{.lrtest}$
			Non-smoker	Former smoker	Current smoker	
Overall survival	rs56963355	*TERT*	1.27 (0.98-1.64)	0.98 (0.75-1.22)	0.48 (0.12-0.84)	0.024
	rs2975842	*TERF1*	0.86 (0.77-0.96)	1.03 (0.93-1.14)	1.06 (0.86-1.27)	0.031
	rs2409652	*PINX1*	0.90 (0.79-1.03)	1.04 (0.92-1.16)	1.23 (0.97-1.49)	0.036
	rs1760899	*TEP1*	0.85 (0.71-1.01)	1.09 (0.93-1.25)	0.64 (0.41-0.87)	0.007
CRC-specific survival	rs56963355	*TERT*	1.34 (0.99-1.80)	0.36 (0.64-1.44)	1.04 (0.73-1.34)	0.031
	rs6420019	*TERT*	1.23 (0.98-1.54)	1.49 (0.45-1.63)	0.85 (0.85-1.22)	0.013
	rs13259648	*TERF2*	0.74 (0.51-1.09)	1.08 (0.40-1.67)	1.49 (0.61-1.46)	0.014
	rs2409652	*PINX1*	0.9 (0.77-1.05)	1.30 (0.67-1.39)	1.08 (0.89-1.18)	0.048
	rs13259648	*PINX1*	1.27 (1.07-1.50)	0.78 (0.75-1.32)	0.99 (0.87-1.19)	0.024
	rs35656875	*TNKS*	0.63 (0.42-0.98)	1.28 (0.09-1.97)	1.43 (0.51-1.56)	0.015
	rs1760901	*TEP1*	0.80 (0.65-0.99)	0.48 (0.78-1.29)	1.12 (0.83-1.24)	0.002
	rs6979	*ACD*	0.96 (0.83-1.11)	0.69 (0.84-1.23)	1.07 (0.89-1.18)	0.023

Abbreviation: HR: hazard ratio; CIs: confidence intervals; $P_{.lrtest:}$ P-value of likelihood ratio test

[a] adjustment for age at diagnosis, sex, study center and the first three principal components (pc)

Table S3 Hazard ratio of the association between SNPs involved in telomere maintenance and survival — stratified by smoking pack-years (25-44 vs. 0) [+]

Outcome	SNP	gene	HR[a] (95% CIs), smoking pack-years		$P_{.lrtest}$	adj. P-value [b]
			0	25-44		
Overall survival	rs2975842	*TERF1*	0.86 (0.77-0.96)	1.15 (0.94-1.36)	0.009	0.054
	rs35431644	*TNKS*	0.84 (0.63-1.13)	1.71 (0.99-2.42)	0.01	0.8
CRC-specific survival	rs73202875	*TERF1*	1.04 (0.89-1.20)	0.73 (0.52-0.95)	0.035	0.21
	rs153045	*TERF2*	1.10 (0.95-1.28)	0.73 (0.51-0.94)	0.013	0.182
	rs118104716	*TERF2IP*	0.76 (0.52-1.12)	1.42 (0.78-2.06)	0.044	0.352
	rs67456872	*TNKS*	0.62 (0.40-0.97)	2.17 (0.76-3.57)	0.003	0.24
	rs56343177	*TEP1*	0.87 (0.71-1.08)	1.32 (0.93-1.71)	0.028	>0.99

Abbreviation: HR: hazard ratio; CIs: confidence intervals; $P_{.lrtest:}$ P-value of likelihood ratio test

[+] only the most significant SNP within gene was presented

[a] adjustment for age at diagnosis, sex, study center and the first three principal components (pc)

[b] p-value is adjusted using Bonferroni methods

Table S4 Hazard ratio of the association between SNPs involved in telomere maintenance and survival — stratified by smoking pack-years (>0-12 vs. 0) [+]

Outcome	SNP	gene	HR[a] (95% CIs), smoking pack-years		$P_{.lrtest}$	adj. P-value [b]
			0	≥45		
Overall survival	rs2975842	*TERF1*	0.86 (0.77-0.96)	1.12 (0.89-1.35)	0.027	0.162
	rs35656875	*TNKS*	0.69 (0.49-0.99)	2.32 (0.96-3.69)	0.001	0.08
	rs78489201	*TNKS1BP1*	0.80 (0.62-1.05)	1.32 (0.81-1.82)	0.043	0.645
	rs1539041	*TNKS2*	0.95 (0.84-1.07)	1.23 (0.97-1.49)	0.036	0.36
CRC-specific survival	rs117840977	*TERF1*	1.18 (0.88-1.57)	0.60 (0.22-0.98)	0.044	0.264
	rs35656875	*TNKS*	0.64 (0.41-0.99)	2.62 (0.89-4.36)	0.001	0.08
	rs1539041	*TNKS2*	0.94 (0.81-1.08)	1.33 (1.00-1.66)	0.017	0.255

Abbreviation: HR: hazard ratio; CIs: confidence intervals; $P_{.lrtest:}$ P-value of likelihood ratio test

[+] only the most significant SNP within gene was presented

[a] adjustment for age at diagnosis, sex, study center and the first three principal components (pc)

[b] p-value is adjusted using Bonferroni methods

Table S5. Hazard ratio of the association sex, smoking and overall survival and CRC-specific survival

	Overall survival		CRC-specific survival	
	HR (95% CIs)	P-value	HR (95% CIs)	P-value
Male [a]	1.18 (1.03-1.04)	0.023*	1.02 (1.00-1.02)	0.863
Ever smoking [b]	1.22 (1.10-1.35)	9.64×10^{-5}***	1.09 (0.97-1.24)	0.147
Smoking status [b]		2.35×10^{-6}***		0.267
current smoker	1.51 (1.29-1.78)		1.16 (0.95-1.43)	
former smoker	1.16 (1.05-1.29)		1.08 (0.95-1.23)	
Smoking pack-years [b]		6.55×10^{-10}***		0.035*
1-11	1.10 (0.93-1.29)		1.08 (0.89-1.32)	
12-24	1.04 (0.88-1.22)		0.96 (0.79-1.17)	
25-44	1.35 (1.16-1.57)		1.22 (1.01-1.48)	
≥45	1.62 (1.41-1.87)		1.28 (1.06-1.54)	

Abbreviation: HR: hazard ratio; CIs: confidence intervals

[a] adjustment for age at diagnosis and study center; females as reference group

[b] adjustment for age at diagnosis, sex and study center; never smokers (0 pack-years) as reference group

* P-value < 0.05, ** P-value < 0.01, *** P-value < 0.001,

3. Power calculation

Study power was calculated by using R-package *survSNP: power calculation for SNP studies with censored outcome*. (22) There are 4896 study subjects, and 1681 any-cause deaths and 1098 CRC-specific deaths. The all-cause death rate was 41.04% and the CRC-specific death rate was 22.43%. MAF was set to scale continuous in the range to generate power-MAF plot. The effect sizes were assumed as 0.8, 1.2 and 1.4, and the median in the population was 1 unit of time. The asymptotic power was checked at the two-sided significance level (α) = 0.05/ M_{eff_G}. After power calculation, SNPs with MAF <5% will be excluded in the analysis since underpower. Results were shown in **Figure S1**, **Figure S2** and **Table S6**.

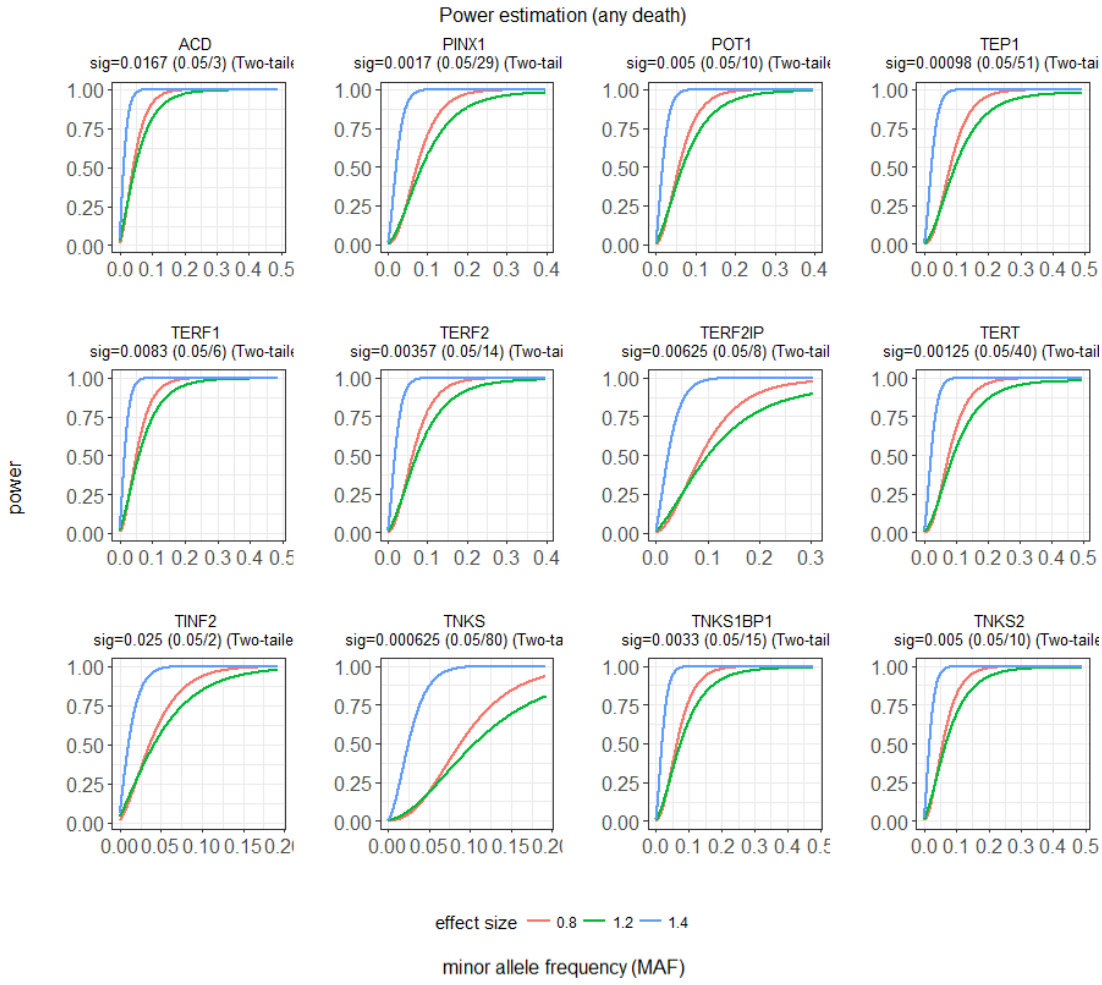

Figure S1. Power calculation for all-cause death

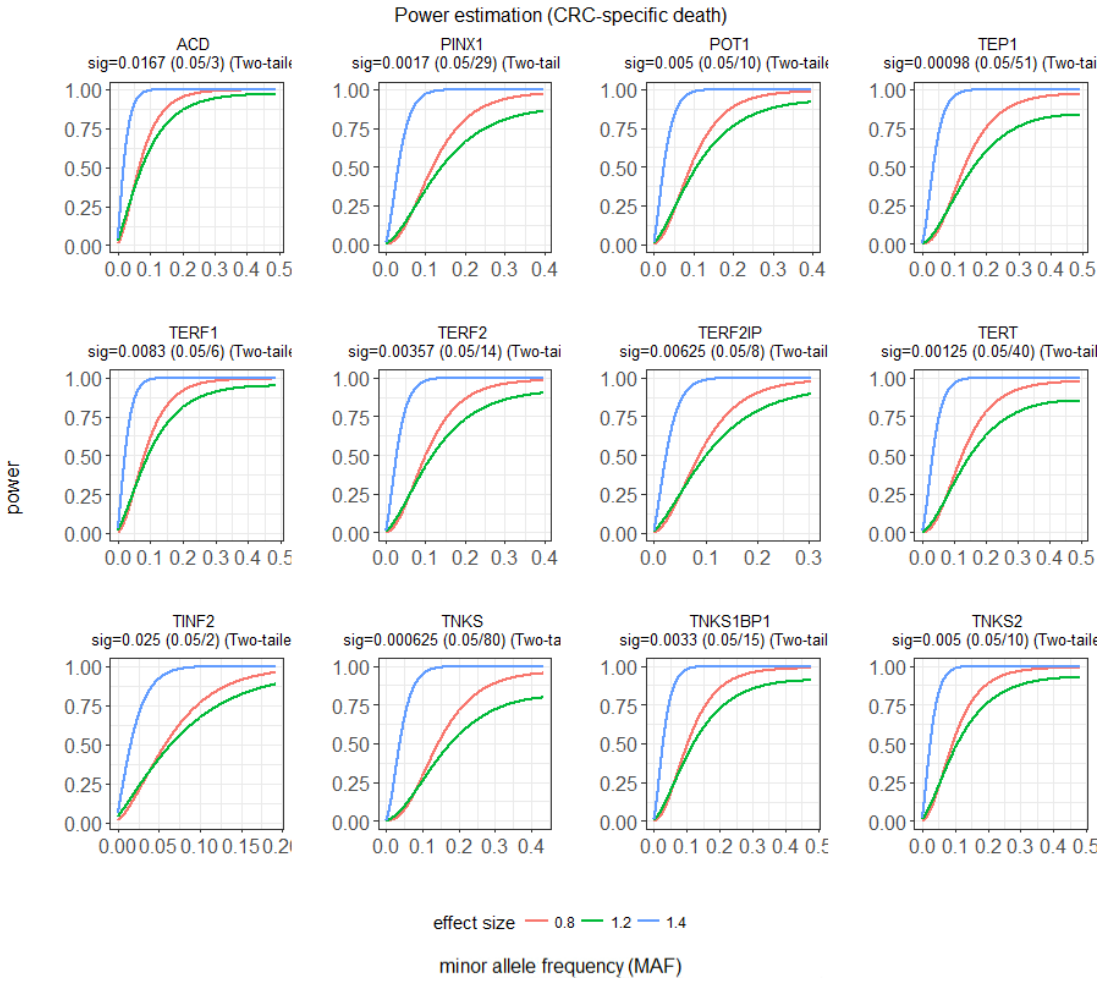

Figure S2. Power calculation for CEC-specific death

Table S6. Power calculation for overall survival and CRC-specific survival

		Overall survival			CRC-specific survival		
Gene	Effect size	0.8	1.2	1.4	0.8	1.2	1.4
	MAF*						
ACD	0.051	60.2%	52.3%	97.9%	37.5%	36.0%	90.4%
	0.100	90.9%	80.9%	>99.9%	71.0%	61.8%	99.4%
PINX1	0.050	28.8%	26.6%	91.8%	21.9%	22.9%	82.8%
	0.234	98.6%	92.2%	>99.9%	93.0%	81.8%	>99.9%
POT1	0.050	43.8%	38.6%	95.8%	22.9%	23.7%	84.0%
	0.234	>99.9%	99.1%	>99.9%	98.8%	92.0%	>99.9%
TEP1	0.051	23.7%	22.5%	90.1%	22.4%	23.4%	83.4%
	0.167	91.2%	78.0%	>99.9%	82.0%	69.4%	99.9%
TERF1	0.053	52.6%	45.7%	97.3%	23.8%	24.4%	84.9%
	0.308	99.9%	98.8%	>99.9%	97.2%	88.5%	>99.9%
TERF2	0.061	48.3%	41.5%	97.5%	28.9%	28.3%	89.4%
	0.268	99.7%	96.6%	>99.9%	95.5%	85.5%	>99.9%
TERF2IP	0.050	45.2%	39.8%	95.8%	21.8%	22.9%	82.7%
	0.138	94.3%	84.6%	>99.9%	72.8%	61.2%	99.7%
TERT	0.053	27.3%	25.3%	91.8%	23.3%	24.0%	84.4%
	0.267	99.1%	93.5%	>99.9%	95.4%	85.3%	>99.9%
TINF2	0.144	98.6%	93.8%	>99.9%	74.9%	63.0%	99.8%
	0.168	99.7%	97.5%	>99.9%	82.2%	69.6%	99.9%
TNKS	0.050	18.8%	18.5%	87.2%	21.6%	22.8%	82.5%
	0.172	89.9%	75.6%	>99.9%	83.1%	70.6%	99.9%
TNKS1BP1	0.051	36.7%	33.0%	94.1%	21.9%	23.0%	82.8%
	0.263	99.6%	96.2%	>99.9%	95.2%	85.0%	>99.9%
TNKS2	0.060	52.5%	45.0%	>99.9%	28.6%	28.0%	>99.9%
	0.182	98.0%	91.3%	99.0%	85.5%	72.9%	92.7%
TERC	0.050	51.7%	45.2%	96.8%	29.8%	29.6%	87.1%
	0.259	99.9%	98.2%	>99.9%	96.9%	89.2%	99.9%

Abbreviation: MAF, minor allele frequency; CRC, colorectal cancer

* Within each gene, minimum (above) and median (below) MAF

References:

1. Slattery ML, Berry TD, Potter J, Caan B. Diet diversity, diet composition, and risk of colon cancer (United States). Cancer Causes Control CCC. 1997;8:872–82.

2. Slattery ML, Friedman GD, Potter JD, Edwards S, Caan BJ, Samowitz W. A description of age, sex, and site distributions of colon carcinoma in three geographic areas. Cancer. 1996;78:1666–70.

3. Slattery ML, Potter J, Caan B, Edwards S, Coates A, Ma KN, et al. Energy balance and colon cancer--beyond physical activity. Cancer Res. 1997;57:75–80.

4. Rimm EB, Stampfer MJ, Colditz GA, Chute CG, Litin LB, Willett WC. Validity of self-reported waist and hip circumferences in men and women. Epidemiol Camb Mass. 1990;1:466–73.

5. Belanger CF, Hennekens CH, Rosner B, Speizer FE. The nurses' health study. Am J Nurs. 1978;78:1039–40.

6. Hennekens CH, Eberlein K. A randomized trial of aspirin and beta-carotene among U.S. physicians. Prev Med. 1985;14:165–8.

7. Christen WG, Gaziano JM, Hennekens CH. Design of Physicians' Health Study II--a randomized trial of beta-carotene, vitamins E and C, and multivitamins, in prevention of cancer, cardiovascular disease, and eye disease, and review of results of completed trials. Ann Epidemiol. 2000;10:125–34.

8. Gohagan JK, Prorok PC, Hayes RB, Kramer BS, Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial Project Team. The Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial of the National Cancer Institute: history, organization, and status. Control Clin Trials. 2000;21:251S-272S.

9. Prorok PC, Andriole GL, Bresalier RS, Buys SS, Chia D, Crawford ED, et al. Design of the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial. Control Clin Trials. 2000;21:273S-309S.

10. National Cancer Institute [Internet]. Cancer Genetic Markers of Susceptibility (CGEMS) data website. Available from: http://cgems.cancer.gov/data_access.html. CGEMS Data accessed October 5, 2009.

11. Yeager M, Chatterjee N, Ciampa J, Jacobs KB, Gonzalez-Bosquet J, Hayes RB, et al. Identification of a new prostate cancer susceptibility locus on chromosome 8q24. Nat Genet. 2009;41:1055–7.

12. Landi MT, Chatterjee N, Yu K, Goldin LR, Goldstein AM, Rotunno M, 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;85:679–91.

13. Newcomb PA, Zheng Y, Chia VM, Morimoto LM, Doria-Rose VP, Templeton A, et al. Estrogen plus progestin use, microsatellite instability, and the risk of colorectal cancer in women. Cancer Res. 2007;67:7534–9.

14. White E, Patterson RE, Kristal AR, Thornquist M, King I, Shattuck AL, et al. VITamins And Lifestyle cohort study: study design and characteristics of supplement users. Am J Epidemiol. 2004;159:83–93.

15. Hays J, Hunt JR, Hubbell FA, Anderson GL, Limacher M, Allen C, et al. The Women's Health Initiative recruitment methods and results. Ann Epidemiol. 2003;13:S18-77.

16. Design of the Women's Health Initiative clinical trial and observational study. The Women's Health Initiative Study Group. Control Clin Trials. 1998;19:61–109.

17. Bergstralh EJ, Kosanke JL. Computerized matching of cases to controls. Rochester (MN): Department of Health Sciences Research, Mayo Clinic; 1995. Technical Report No.: 56..

18. Sever ML, Salo PM, Haynes AK, Zeldin DC. Inner-city environments and mitigation of cockroach allergen. Am J Prev Med. 2011;41:S55-56.

19. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006;38:904–9.

20. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. Genet Epidemiol. 2010;34:816–34.

21. Jiao S, Hsu L, Hutter CM, Peters U. The use of imputed values in the meta-analysis of genome-wide association studies. Genet Epidemiol. 2011;35:597–605.

22. Owzar K, Li Z, Cox N, Jung S-H. Power and Sample Size Calculations for SNP Association Studies with Censored Time-to-Event Outcomes. Genet Epidemiol. 2012;36:538–48.