SHORT REPORT

Smoking and colorectal cancer survival in relation to tumor LINE-1 methylation levels: a prospective cohort study

Junko Kishikawa1,2*, Tomotaka Ugai1,3*, Kenji Fujiyoshi1,4†, Yang Chen1,5†, Koichiro Haruki1,7†, Li Liu1, Kota Arima1, Naohiko Akimoto1, Tsuyoshi Hamada1, Kentaro Inamura1, Keisuke Kosumi1, Tyler S. Twombly1, Shanshan Shi1, Mai Chan Lau1, Chunxia Du1, Peilong Li1, Chunquang Guo1, Juha P. Väyrynen1,6,7, Sara A. Väyrynen1, Mingyang Song5,9,10, Xuehong Zhang1,11, Guanghai Dai1,2, Marios Giannakis6,13,14, David A. Drew8,9, Andrew T. Chan8,9,11,15, Charles S. Fuchs16,17,18,19, Jeffrey A. Meyerhardt7†, Reiko Nishihara1,2,10,20†, Jonathan A. Nowak1†, Shuji Ogino1,3,13,21† and Kana Wu10†

Abstract

Background: Carcinogens in cigarette smoke may cause aberrant epigenomic changes. The hypomethylation of long interspersed nucleotide element-1 (LINE-1) in colorectal carcinoma has been associated with genomic instability and worse clinical outcome. We hypothesized that the association between smoking behavior and colorectal cancer mortality might be stronger in tumors with lower LINE-1 methylation levels.

Findings: To test our hypothesis, we examined the interaction of tumor LINE-1 methylation levels and smoking status at diagnosis using data of 1208 cases among 4420 incident colorectal cancer cases that were ascertained in two prospective cohort studies. We conducted multivariable Cox proportional hazards regression analyses, using inverse probability weighting with covariate data of the 4420 cases to control for potential confounders and selection bias due to data availability. The prognostic association of smoking status at diagnosis differed by tumor LINE-1 methylation levels (Pinteraction = 0.050 for overall mortality and 0.017 for colorectal cancer-specific mortality; with an alpha level of 0.005). In cases with <60% LINE-1 methylation, current smoking (vs. never smoking) was associated with worse overall mortality (multivariable hazard ratio, 1.80; 95% confidence interval, 1.19–2.73). In contrast, smoking status was not associated with mortality in cases with ≥60% LINE-1 methylation.

Conclusions: Our findings suggest that the association between smoking status and mortality is stronger in colorectal cancer patients with lower tumor LINE-1 methylation levels. These results warrant further investigation into an interactive role of smoking and aberrant DNA methylation in colorectal cancer progression.
Introduction
Cigarette smoke is an established risk factor for colorectal cancer incidence and mortality [1–3]. Cigarette smoke contains hundreds of carcinogens, some of which can cause epigenetic alterations [4]. While there is strong evidence linking smoking to epigenetic changes, including findings from Epigenome-Wide Association Studies, less is known about epigenetic changes in former smokers, particularly long-time quitters [5, 6].

Accumulating evidence suggests that DNA hypomethylation may play an important role in colorectal cancer progression [7, 8]. Long interspersed nucleotide element-1 (LINE-1) hypomethylation is a surrogate marker for genome-wide DNA hypomethylation [9], and LINE-1 hypomethylated colorectal cancer has been associated with worse prognosis and non-response to certain chemotherapies, suggesting potential utilization of LINE-1 hypomethylation as a prognostic biomarker [10, 11]. Moreover, LINE-1 hypomethylation is an emerging biomarker for early-onset colorectal cancer diagnosed before age 50 [12], which has shown increasing incidence in many parts of the world since the 1980s [13].

To test whether the association of smoking status at diagnosis with colorectal cancer mortality might differ by LINE-1 methylation levels in tumors, we leveraged data from two large prospective cohorts with 4420 incident colorectal cancer cases, including 1208 cases with available tumor tissue data.

Methods
Study population and design
The Nurses’ Health Study (NHS, N = 121,701) was established in 1976 and the Health Professionals Follow-up Study (HPFS, N = 51,529) was established in 1986 [14, 15]. Self-administered questionnaires were mailed to participants at baseline and then biennially to update smoking status, lifestyle, and medical history. Semi-quantitative food frequency questionnaires were administered every 4 years to assess participants’ diet.

Deaths were identified through next-of-kin reports or the National Death Index, and cause of death was determined by study physicians after a review of the medical records or death certificates. A pathologist (S.O.), blinded to other information, conducted a centralized review of hematoxylin and eosin (H&E)-stained tissue sections of all colorectal cancer cases and recorded pathological features including tumor differentiation. The study protocol was approved by the institutional review boards of the Brigham and Women’s Hospital and the Harvard T. H. Chan School of Public Health, and those of participating registries as required. We also obtained signed consents from patients (or next-of-kin for deceased patients) to use CRC tissue specimens for molecular pathological analyses.

Assessment of smoking behavior at diagnosis
Detailed information on smoking was obtained as previously described [1, 2, 15]. Smoking status was divided into three categories (never smoking, past smoking, and current smoking at the time of diagnosis). As most current smokers (86%) had a smoking history of ≥20 pack-years at diagnosis, we did not examine associations by pack-years of smoking in current smokers. Of the 599 past smokers, 285 (47.6%) had a smoking history of 1–19 pack-years and 314 cases (52.4%) had a history of pack-years of ≥20. Past smokers who quit ≥10 years prior to diagnosis were more likely to have a smoking history of 1–19 pack-years than those who quit <10 years prior to diagnosis (201 out of 322 or 62% vs. 84 out of 277 or 30%); therefore, for past smokers, we did not further stratify by time since quitting smoking (Additional file 1: Supplementary Table 1). Smoking status, diet, and lifestyle at the time of diagnosis were defined using participants’ most recent available questionnaire prior to the diagnosis of cancer.

Analyses of LINE-1 methylation analysis, microsatellite instability (MSI), CpG island methylator phenotype (CIMP), and KRAS, BRAF, and PIK3CA mutations
More details on our methods and pyrosequencing results were described in our previous publications [11, 16]. In brief, DNA was extracted from formalin-fixed paraffin-embedded tissue blocks, focusing on tumor areas only. We performed bisulfite DNA treatment, polymerase chain reaction (PCR), and pyrosequencing using the PyroMark kit (Qiagen), and quantified LINE-1 methylation levels by amplifying a region of the LINE-1 element (position 305 to 331 in accession No. X58075) which includes 4 CpG sites. We used the average of the proportions of C nucleotides at the 4 CpG sites (a scale of 0 to 100) as the LINE-1 methylation levels of each case. The LINE-1 methylation level showed a normal distribution [16] and was used both as a continuous variable (scale 0–100%) and a categorical variable (“high” (≥68% methylation), “intermediate” (≥60% and <68% methylation), and “low” (<60% methylation)). The precision of pyrosequencing assay has been validated in our
previous study using approximately 500 cancer cells from 5 anonymized colorectal cancer cases which were collected by laser capture microdissection [17].

MSI analysis was carried out utilizing a panel of 10 microsatellite markers, as previously described [18]. MSI-high was defined as instability in ≥30% of the markers. We quantified DNA methylation in eight CpG island methylator phenotype (CIMP)-specific promoters (CACNA1G, CDKN2A, CRABP1, IGF2, MLH1, NEUROG1, RUNX3, and SOCS1) by using bisulfite DNA treatment and real-time PCR (MethyLight) as previously described [19, 20]. PCR and pyrosequencing targeted for KRAS (codons 12, 13, 61, and 146) [21, 22], BRAF (codon 600) [18], and PIK3CA (exons 9 and 20) were performed as previously described [23].

**Statistical analyses**

Our primary hypothesis testing was an assessment of a statistical interaction between smoking status at diagnosis (ordinal categories; never smoker (0), past smoker with 1–19 packyears (1), past smoker with ≥20 packyears (2), and current smoker (3)) and tumor LINE-1 methylation levels (continuous) using Cox proportional hazards regression. We utilized the Wald test to assess the statistical significance of that interaction and reported the \( P \) value as \( P_{interaction} \). Our main endpoints were all-cause and colorectal cancer-specific mortality. Survival time was defined as the time from colorectal cancer diagnosis until death or the end of follow-up, whichever came first (January 1, 2016, for the HPS; May 31, 2016, for the NHS). For the analysis of colorectal cancer-specific mortality, deaths from other causes were censored. We calculated hazard ratio (HR) and its 95% confidence interval (CI) using re-parameterization of the interaction term in a single regression model. The trend test was conducted using the ordinal smoking variable.

To reduce potential selection bias due to tumor tissue data availability, we applied the inverse probability weighting (IPW) method using all 4420 cases as reported previously [24, 25].

We used the multivariable IPW-adjusted Cox proportional hazards regression models to adjust for potential confounders, which initially included age at diagnosis, year of diagnosis of cancer family history of colorectal cancer in first degree relatives, body mass index at diagnosis (BMI), alcohol consumption at diagnosis, empirical dietary inflammatory pattern (EDIP) score at diagnosis, dietary fiber intake at diagnosis, folic acid intake at diagnosis, regular aspirin use at diagnosis, physical activity at diagnosis, tumor location and differentiation, MSI and CIMP status, and Kras, BRAF, and PIK3CA mutations. We conducted a backward elimination with a threshold of \( P = 0.05 \) to select variables for the final models. The proportionality of hazards assumption was generally satisfied after assessment of Schoenfeld residual plots and including interaction terms of smoking and survival time to the multivariable models. All \( P \) values were two-sided and a \( P < 0.005 \) was considered statistically significant as recommended by the expert panel [26].

**Results**

**Patients’ characteristics**

With a follow-up to 2012, we documented 4420 incident colorectal cancer cases including 1208 patients with available data on both smoking information at diagnosis and tumor LINE-1 methylation level (Additional file 1: Supplementary Table 2). After a median follow-up time of 16 years (interquartile range 11.9–20.3 years) for censored cases, 776 all-cause deaths including 343 colorectal cancer-specific deaths were identified.

**Smoking status and mortality**

Compared with never smokers, HRs of past smokers with <19 packyears, past smokers with ≥20 packyears, and current smokers were 0.81 (0.65–1.00), 1.01 (0.85–1.21), and 1.31 (0.99–1.71) for all-cause mortality and 0.79 (0.58–1.08), 0.86 (0.65–1.13), and 1.07 (0.75–1.53) for colorectal cancer-specific mortality, respectively (Additional file 1: Supplementary Table 3).

**Tumor LINE-1 methylation levels and mortality**

Colorectal cancer cases who developed tumors with high LINE-1 methylation level tumors (≥68%) appeared to have a lower risk of colorectal cancer-specific mortality (HR 0.74, 95% CI 0.55–0.99) than those with lower (<60%) LINE-1 methylation levels. However, the lower risk in cases with high LINE-1 methylation level tumors was not seen for all-cause mortality (Additional file 1: Supplementary Table 4).

**Smoking and mortality in relation to tumor LINE-1 methylation levels**

In our primary hypothesis testing, the association between smoking status and mortality appeared to be stronger in cases with low levels of LINE-1 methylation tumors than in those with high levels of LINE-1 methylation (\( P_{interaction} = 0.050 \) for all-cause mortality, \( P_{interaction} = 0.017 \) for colorectal cancer-specific mortality; Table 1). Among cases with low LINE-1 methylation tumors (<60%), HRs of current smoking relative to never smoking were 1.55 (95% CI, 0.94–2.56) for colorectal cancer mortality and 1.80 (95% CI, 1.19–2.73) for all-cause mortality.
On the other hand, among cases with high levels of LINE-1 methylation (≥68%), corresponding HRs (current vs. never smoking) were 0.93 (95% CI, 0.50–1.73) for colorectal cancer-specific mortality and 1.33 (95% CI, 0.85–2.08) for all-cause mortality. Results did not change substantially when we repeated analyses without using IPW adjustment, or restricted analyses to cases with stage I–III colorectal cancers (Additional files 1: Supplementary Tables 5 and 6).

### Discussion

Previous studies have suggested that smoking may be associated with worse survival in colorectal cancer, perhaps in certain tumor molecular subtypes though findings have not been consistent [27–30]. In contrast to those tumor markers (except for MSI status) in the previous studies [27–29], tumor LINE-1 hypomethylation has been consistently shown to be a strong prognostic indicator in various cancer types including colorectal cancer [12, 31, 

### Table 1

Smoking status at diagnosis\(^a\) and colorectal cancer mortality stratified by tumor LINE-1 methylation levels in the Nurses’ Health Study and Health Professionals Follow-up Study

| LINE-1 methylation level | No. of cases | No. of events | Colorectal cancer-specific mortality | Multivariable HR\(^d\) (95% CI) | Overall mortality | Multivariable HR\(^e\) (95% CI) |
|--------------------------|--------------|--------------|-------------------------------------|-------------------------------|------------------|-------------------------------|
| < 60% (low)              |              |              |                                     |                               |                  |                               |
| Never smoker             | 167          | 53           | 1 (referent)                        |                               | 1 (referent)     |                               |
| Past smoker              |              |              |                                     |                               |                  |                               |
| < 19 packyears           | 94           | 35           | 1.05 (0.66–1.68)                    | 56                            | 0.91 (0.64–1.31) |
| ≥ 20 packyears           | 103          | 37           | 1.16 (0.75–1.81)                    | 75                            | 1.17 (0.84–1.62) |
| Current smoker           | 37           | 19           | 1.55 (0.94–2.56)                    | 32                            | 1.80 (1.19–2.73) |
| \(P\text{\_trend}\)\(^d\) |              |              | 0.14                                |                               |                  | 0.024                         |
| 60–67.9% (intermediate)  |              |              |                                     |                               |                  |                               |
| Never smoker             | 161          | 43           | 1 (referent)                        | 101                           | 1 (referent)     |                               |
| Past smoker              |              |              |                                     |                               |                  |                               |
| < 19 packyears           | 98           | 20           | 0.54 (0.30–0.95)                    | 55                            | 0.66 (0.47–0.93) |
| ≥ 20 packyears           | 100          | 28           | 0.68 (0.42–1.10)                    | 77                            | 0.95 (0.70–1.28) |
| Current smoker           | 44           | 14           | 0.77 (0.37–1.59)                    | 35                            | 1.01 (0.64–1.60) |
| \(P\text{\_trend}\)\(^d\) |              |              | 0.24                                |                               |                  | 0.91                          |
| ≥ 68% (high)             |              |              |                                     |                               |                  |                               |
| Never smoker             | 156          | 44           | 1 (referent)                        | 93                            | 1 (referent)     |                               |
| Past smoker              |              |              |                                     |                               |                  |                               |
| < 19 packyears           | 93           | 16           | 0.83 (0.44–1.57)                    | 44                            | 0.89 (0.60–1.32) |
| ≥ 20 packyears           | 111          | 21           | 0.77 (0.47–1.24)                    | 68                            | 0.92 (0.68–1.25) |
| Current smoker           | 44           | 13           | 0.93 (0.50–1.73)                    | 30                            | 1.33 (0.85–2.08) |
| \(P\text{\_trend}\)\(^d\) |              |              | 0.55                                |                               | 0.47             |                               |
| \(P\text{\_interaction}\)\(^d\) | | | 0.017 | 0.050 | |

\(a\) As most current smokers (86%) had a smoking history of ≥20 packyears at diagnosis, we did not examine associations by packyears of smoking in current smokers. Of the 599 past smokers, 285 (47.6%) had a smoking history of 1–19 packyears and 314 cases (52.4%) had a history of ≥20 packyears. Past smokers who quit ≥10 years prior to diagnosis were more likely to have a smoking history of 1–19 packyears than those who quit <10 years prior to diagnosis (201 out of 322 or 62% vs. 84 out of 277 or 30%); therefore, for past smokers, we did not further stratify by time since quitting smoking (Additional file 1: Supplementary Table 1).

\(b\) Inverse probability weighting was applied to reduce selection bias due to tissue data availability.

\(c\) The multivariable Cox regression model initially included age, sex, year of diagnosis, family history of colorectal cancer, body mass index, alcohol consumption at diagnosis, empirical dietary inflammatory pattern (EDIP) score at diagnosis, dietary fiber intake at diagnosis, folate intake at diagnosis, regular aspirin intake at diagnosis, physical activity status at diagnosis, tumor location, tumor differentiation, microsatellite instability, CpG island methylator phenotype, KRAS mutation, BRAF mutation, and PIK3CA mutation. A backward elimination with a threshold of \(P = 0.05\) was used to select variables in the final models.

\(d\) The trend test was conducted using an ordinal smoking variable (never smoker [0], past smoker with 1–19 packyears [1], past smoker with ≥20 packyears [2], and current smoker [3]).

\(e\) \(P\text{\_interaction}\) was calculated using the Wald test for the cross-product term of smoking status at diagnosis (ordinal; never smoker, past smoker with <19 packyears, ≥20 packyears, current smoker) and LINE-1 methylation level (continuous) in Cox regression models.

Abbreviations: CI confidence interval, HR hazard ratio, LINE-1 long interspersed nucleotide element-1

---

[Additional files 1: Supplementary Tables 5 and 6]
Our findings suggest that the positive association between smoking status at diagnosis and mortality may be more pronounced in cases with low LINE-1 methylation levels than in those with high LINE-1 methylation levels. Experimental studies have shown that DNMT1 (DNA methyltransferase 1) mutation can lead to global DNA hypomethylation [7] and that global DNA hypomethylation promotes tumor development through chromosomal instability including loss of heterozygosity of TP53, which can cause cell cycle arrest for DNA repair or apoptosis of damaged cells [33]. Assuming that tumors with high genomic instability are more prone to somatic mutation induced by exogenous factors such as smoking, it is possible that cases who later on developed tumors with LINE-1 hypomethylation may have been more susceptible to the mutagenic effects of smoking than those with high LINE-1 methylation tumors. Our findings suggest that the stronger association between smoking and mortality observed in LINE-1 hypomethylated tumors may be, at least in part, explained by a higher accumulation of genomic instability over time. In addition, tumor LINE-1 hypomethylation has been associated with lower levels of T cell immune response to colorectal cancer, suggesting its immunosuppressive effect [34]. Furthermore, smoking has been associated with the incidence of colorectal cancer subtypes containing fewer counts of T cells and macrophages, implying its suppressive effect on effector immune cells [35, 36].

Together with these previous findings, our current results may suggest that smoking status and tumor LINE-1 hypomethylation interact and jointly influence the tumor-immune interaction, leading to a stronger prognostic role of smoking status for tumors with LINE-1 hypomethylation.

Strength and limitations
One major strength of this study was our molecular pathological epidemiology [37–39] database of colorectal cancer cases with the availability of diet and lifestyle information that has been prospectively and repeatedly collected. This rich database enabled us to examine the prognostic interaction between smoking behavior and tumor LINE-1 methylation levels while adjusting for multiple potential confounders [40, 41] and selection bias due to tumor molecular data availability [25].

There are several limitations to our study. First, we did not examine the effect of postdiagnosis smoking status because 40% of current smokers (44 out of 110) quit smoking after diagnosis and the sample size of postdiagnosis current smokers was small. Second, data on cancer treatment were limited in this dataset. However, it is unlikely that the ratio of patients who underwent chemotherapy differed substantially according to smoking status. Additionally, treatment protocols for colorectal cancers are generally similar across the USA and adjusting for the AJCC stage should have limited potential confounding due to treatment. Third, data on cancer recurrence were unavailable. However, given a follow-up of >10 years, colorectal cancer-specific mortality can be considered as a reasonable measure of colorectal cancer outcome. Fourth, our main results did not meet our stringent multiple comparison-adjusted significance level of P < 0.005. However, we selected all the risk factors and statistical comparisons on the basis of previous data and certain hypotheses and interpreted our results prioritizing biological plausibility, coherence, and consistency rather than only statistical significance. Fifth, there is evidence that LINE-1 hypomethylation is inversely associated with MSI-high, CIMP-high, and BRAF-mutated CRC [11, 16]. While we adjusted for these molecular markers in our multivariable models, due to limited sample size, we were not able to further stratify by these markers. Therefore, our findings warrant additional investigation in future larger-sized studies with sufficient power for these stratified analyses. Sixth, we cannot exclude the possibility of biases related to tumor heterogeneity and contaminated normal cells. In addition, a previous study reported cell-type heterogeneity in LINE-1 methylation levels [42], which might affect our results. However, an experienced pathologist, Dr. Shuji Ogino, carefully reviewed H&E-stained slides of all cases and identified tumor areas in each section, which minimized the possibility of these biases.

Future prospects
Our study has shown a prognostic interaction between smoking and tumor LINE-1 methylation levels measured by the bisulfite-PCR-pyrosequencing method. We used the average of the proportions of C nucleotides at the 4 CpG sites as LINE-1 methylation levels, but DNA methylation may vary in specific repetitive elements of genomes. Recently, several novel approaches have been developed to explore detailed epigenetic profiling. Bock et al. have shown that locus-specific DNA methylation assays in combination with machine learning algorithms can predict global DNA methylation levels more accurately [43]. Subsequently, Zhang et al. reported that a random forest algorithm could accurately predict genome-wide repetitive element methylation using microarray data [44]. Furthermore, nanopore sequencing has enabled us to conduct a direct and real-time analysis of long DNA fragments electronically, which leads to the elimination
of amplification bias and efficient assembly, compared to short-read sequencing [45]. These methods can be used in future studies for accurately measuring DNA methylation levels.

Conclusion
Our findings suggest that the association of smoking status at diagnosis with colorectal mortality may be stronger in cases with low LINE-1 methylation level tumors than in those with intermediate or high LINE-1 methylation level tumors. Considering the need for more accurate CRC prognostication, future larger-sized studies are warranted to confirm our findings and guide further exploration into underlying pathways.

Abbreviations
Cf: Confidence interval; CIMP: CpG island methylator phenotype; H&E: Hematoxylin and eosin; HPFS: Health Professionals Follow-up Study; HR: Hazard ratio; IPW: Inverse probability weighting; LINE-1: Long interspersed nucleotide element-1; METS: Metabolic equivalent task score-hours; MPE: Molecular pathological epidemiology; MSi: Microsatellite instability; NHS: Nurses’ Health Study; PCR: Polymerase chain reaction; SD: Standard deviation.

Supplementary Information
The online version contains supplementary material available at https://doi.org/10.1186/s43682-022-00012-y.

Acknowledgements
We deeply thank hospitals and pathology departments throughout the USA for generously providing us with tissue specimens. In addition, 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, and WY. The authors assume full responsibility for the analyses and interpretation of these data.

Authors’ contributions
K.W., R.N., S.O., Y.C., K.H., K.F., L.L., T.U., K.A., N.A., T.H., K.I., K.K., S.S., C.D., P.L., C.G., J.P.V., S.G., G.D., M.G., E.G., J.K., Y.C., K.H., K.F., L.L., T.U., R.N., and K.W. performed the data analysis and interpretation of these data.

Supplementary Information
The online version contains supplementary material available at https://doi.org/10.1186/s43682-022-00012-y.

Additional file 1: Supplementary Table 1. Frequencies between smoking status and pack-years among past/current smokers.

Availability of data and materials
The datasets generated during and/or analyzed during the current study are not publicly available due to participant confidentiality and privacy concerns. Further information including the procedures to obtain and access data from the Nurses’ Health Studies and Health Professionals Follow-up Study is described at https://www.nurseshealthstudy.org/researchers (contact email: rhaccess@channing.harvard.edu) and https://sites.sph.harvard.edu/hpfs/for-collaborators/.

Declarations
Ethics approval and consent to participate
The study protocol was approved by the institutional review boards of the Brigham and Women’s Hospital and Harvard T.H. Chan School of Public Health, and those of participating registries as required. All participants provided informed consent.

Consent for publication
Not applicable.

Competing interests
A.T.C. previously served as a consultant for Bayer Pharma AG, Pfizer Inc., and Janssen. This study was not funded by Bayer Pharma AG, Pfizer Inc., or Janssen. M.G. receives research funding from Bristol-Myers Squibb and Merck. This study was not funded by Bristol-Myers Squibb or Merck. C.S.F. has been a consultant for Eli Lilly and Co; Entinrix Health Solutions, Inc; Pfizer Inc; Merck & Co.; and Sanofi A.F.; Hoffmann-La Roche Ltd; Genetech, Inc; Merimy Pharmaceuticals, Inc; Dicerna Pharmaceuticals, Inc; Bayer Pharma AG, Agios Pharmaceuticals, Inc; Gilead Sciences, Inc; Five Prime Therapeutics, Inc; and Taiho Pharmaceutical Co., Ltd. C.S.F. is currently employed by Genentech, Roche. J.A.M. has also served as an advisor/consultant to Ignyta, Array Pharmaceutical, and Cota. R.N. is currently employed by Pfizer Inc.; she contributed to this study before she became an employee of Pfizer Inc. This study was not funded by any of these commercial entities. The other authors declare that they have no conflicts of interest.

Use of standardized official symbols
We use HUGO (Human Genome Organisation)-approved official symbols for genes and gene products, including BRAF, CACNA1G, CDKN2A, CRABP1, DNM1, IGF2, KRAS, MLH1, NEUROG1, PIK3CA, RUNX1, SOCS1, and TP53, all of which are described at www.genenames.org. The official symbols are italicized to differentiate from non-italicized colloquial names that are used along with the official symbols. This format enables readers to familiarize themselves with the official symbols for genes and gene products together with common colloquial names.

Author details
1Department of Pathology, Brigham and Women’s Hospital and Harvard Medical School, Program in MPE Molecular Pathological Epidemiology, 221 Longwood Ave., EBRC Room 404, Boston, MA 02215, USA. 2Department of Surgical Oncology, The University of Tokyo, 7-3-1, Hongo, Tokyo, Bunkyo 113-8655, Institute for Cancer Research (to K.W.); Nodal Award (2016-02) from the Dana-Farber Harvard Cancer Center (to S.O.); and by grants from the Project P Fund; The Friends of the Dana-Farber Cancer Institute, Bennett Family Fund, and the Entertainment Industry Foundation through National Colorectal Cancer Research Alliance and SU2C, National Natural Science Foundation of China (81872286) to G.D. Y.C., P.L., and L.L. were supported by a scholarship grant from Chinese Scholarship Council. K.K., K.A., and T.U. were supported by grants from the Overseas Research Fellowship from the Japan Society for the Promotion of Science (JP2017-775 to K.K.; JP2018B60083 to K.A.; JP201960541 to T.U.). M.G. was supported by a Conquer Cancer Foundation of ASCO Career Development Award. K.F., K.H., and T.U. were supported by fellowship grants from the Uehara Memorial Foundation. T.U. was supported by grants from Mishima Kaisen Memorial Foundation and Prevent Cancer Foundation. K.H. was supported by a fellowship from the Mitsubishi Health and Welfare Foundation. S.A.V. was supported by grants from the Finnish Cultural Foundation and Onkon Research Foundation sr. A.T.C. is a Stuart and Suzanne Steele MGH Research Scholar. The content is solely the responsibility of the authors and does not necessarily represent the official views of NIH. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.
Japan. 2Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA. 3Department of Surgery, Kurume University, Kurume, Japan. 4Department of Gastrointestinal Oncology, Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education/Beijing), Peking University Cancer Hospital and Institute, Beijing, China. 5Cancer and Translational Medicine Research Unit, Medical Research Council Oulu, Oulu University Hospital, and University of Oulu, Oulu, Finland. 6Department of Medical Oncology, Dana-Farber Cancer Institute and Harvard Medical School, Boston, MA, USA. 7Clinical and Translational Epidemiology Unit, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA. 8Division of Gastroenterology, Massachusetts General Hospital, Boston, MA, USA. 9Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA, USA. 10Department of Medicine, Brigham and Women’s Hospital and Harvard Medical School, Channing Division of Network Medicine, Boston, MA, USA. 11Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Boston, MA, USA. 12Department of Medicine, Brigham and Women’s Hospital and Harvard Medical School, Channing Division of Network Medicine, Boston, MA, USA. 13Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA, USA. 14Department of Medicine, Brigham and Women’s Hospital and Harvard Medical School, Channing Division of Network Medicine, Boston, MA, USA. 15Department of Nutrition, Dana-Farber Cancer Center, Boston, MA, USA. 16Department of Medicine, Brigham and Women’s Hospital and Harvard Medical School, Dana-Farber Cancer Institute and Harvard Medical School, Boston, MA, USA. 17Department of Medicine, Brigham and Women’s Hospital and Harvard Medical School, Boston, MA, USA. 18Smilow Cancer Hospital, New Haven, CT, USA. 19Genentech, South San Francisco, CA, USA. 20Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA. 21Cancer Immunology and Cancer Epidemiology Programs, Dana-Farber Cancer Center, Boston, MA, USA.

Received: 20 August 2021   Accepted: 7 June 2022

Published online: 20 June 2022

References

1. Giovannucci E, Colditz GA, Stampfer MJ, et al. A prospective study of cigarette smoking and risk of colorectal adenoma and colorectal cancer in U.S. women. J Natl Cancer Inst. 1994;86(3):192–9.
2. Giovannucci E, Rimm EB, Stampfer MJ, et al. A prospective study of cigarette smoking and risk of colorectal adenoma and colorectal cancer in U.S. men. J Natl Cancer Inst. 1994;86(3):183–91.
3. Alves E, Carr PR, Bandury B, et al. Smoking behavior and prognosis after colorectal cancer diagnosis: a pooled analysis of 11 studies. JNCI Cancer Spectr. 2021;5(5):pkab077.
4. Wang TH, Hsia SM, Shih YH, et al. Association of smoking, alcohol use, and betel quid chewing with epigenetic aberrations in cancers. Int J Mol Sci. 2017;18(6):1210.
5. Zellinger S, Kühnel B, Kopp N, et al. Tobacco smoking leads to extensive genome-wide changes in DNA methylation. PLoS ONE. 2013;8(5):e63812.
6. Wan ES, Qiu W, Baccarelli A, et al. DNA hypomethylation contributes to genomic instability and intestinal cancer initiation. Cancer Prev Res (Philadelphia, Pa). 2016;9(7):534–46.
7. Sheaffer KL, Elliott EN, Kaestner KH. DNA hypomethylation and prognosis in colon cancer. J Natl Cancer Inst. 2018;100(23):1734–8.
38. Ogino S, Nowak JA, Hamada T, et al. Insights into pathogenic interactions among environment, host, and tumor at the crossroads of molecular pathology and epidemiology. Annu Rev Pathol. 2019;14:83–103.

39. Mima K, Kosumi K, Baba Y, et al. The microbiome, genetics, and gastrointestinal neoplasms: the evolving field of molecular pathological epidemiology to analyze the tumor-immune-microbiome interaction. Hum Genet. 2021;140(5):725–46.

40. Phipps AI, Passarelli MN, Chan AT, et al. Common genetic variation and survival after colorectal cancer diagnosis: a genome-wide analysis. Carcinogenesis. 2016;37(1):87–95.

41. Liu L, Nishihara R, Qian ZR, et al. Association between inflammatory diet pattern and risk of colorectal carcinoma subtypes classified by immune responses to tumor. Gastroenterology. 2017;153(6):1517–30.e14.

42. Knothe C, Shiratori H, Resch E, et al. Disagreement between two common biomarkers of global DNA methylation. Clin Epigenetics. 2016;8:60.

43. BLUEPRINT consortium. Quantitative comparison of DNA methylation assays for biomarker development and clinical applications. Nat Biotechnol. 2016;34(7):726–37.

44. Zheng Y, Joyce BT, Liu L, et al. Prediction of genome-wide DNA methylation in repetitive elements. Nucleic Acids Res. 2017;45(15):8697–711.

45. Ewing AD, Smits N, Sanchez-Luque FJ, et al. Nanopore sequencing enables comprehensive transposable element epigenomic profiling. Mol Cell. 2020;80(3):913–28.

**Publisher’s Note**

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