BMJ Oncology

Smoking habit and long-term colorectal cancer incidence by exome-wide mutational and neoantigen loads: evidence based on the prospective cohort incident-tumour biobank method

Tsuyoshi Hamada ,[1,2] Tomotaka Ugai ,[1,3] Carino Gurjao,[4,5] Satoko Ugai,[1,3] Xuehong Zhang ,[6,7] Koichiro Haruki ,[1] Yasutoshi Takashima,[1] Naohiko Akimoto,[1] Mai Chan Lau,[1] Kosuke Matsuda,[1] Nobuhiro Nakazawa,[1] Mayu Higashioka,[1] Satoshi Miyahara,[1] Keisuke Kosumi,[1] Yohei Masugi,[1] Li Liu,[1,6,8] Yin Cao,[9,10] Daniel Nevo,[3,11] Molin Wang,[3,11,12] Reiko Nishihara,[1,3,6,11] Sachet A Shukla,[13] Catherine J Wu,[4,5,14] Levi A Garraway,[4,5,14] Jeffrey A Meyerhardt,[4] Edward L Giovannucci ,[3,6] Jonathan A Nowak ,[1] Charles S Fuchs,[15] Andrew T Chan ,[5,10,12,16,17] Mingyang Song,[6,10,16] Marios Giannakis ,[4,5,14] Shuji Ogino [1,3,5,18,19]

To cite: Hamada T, Ugai T, Gurjao C, *et al*. Smoking habit and long-term colorectal cancer incidence by exome-wide mutational and neoantigen loads: evidence based on the prospective cohort incident-tumour biobank method. *BMJ Oncology* 2025;**4**:e000787. doi:10.1136/bmjonc-2025-000787

TH, TU, CG and SU are joint first authors.
ATC, MS, MG and SO are joint senior authors.

Received 26 February 2025
Accepted 21 April 2025

► https://doi.org/10.1136/bmjonc-2025-000855

For numbered affiliations see end of article.

Correspondence to
Dr Tsuyoshi Hamada; hamada-tky@umin.ac.jp and Dr Shuji Ogino; sogino@bwh.harvard.edu

ABSTRACT

Objective To test the hypothesis that the association of smoking with long-term colorectal cancer incidence may be stronger for tumours with higher mutational and neoantigen loads.

Methods and analysis In the Nurses' Health Study (1980–2012) and the Health Professionals Follow-up Study (1986–2012), our novel prospective cohort incident-tumour biobank method (PCIBM) used 3053 incident colorectal carcinoma cases including 752 cases with whole-exome sequencing data. Using the multivariable duplication-method Cox regression model with the inverse probability weighting to adjust for the selection bias due to tissue availability, we assessed a differential association of cigarette smoking with colorectal carcinoma incidence by an exome-wide tumour mutational burden (e-TMB) or neoantigen load.

Results The association of pack-years smoked with colorectal cancer incidence differed by e-TMB ($P_{heterogeneity}$<0.001). Multivariable-adjusted HRs for e-TMB-high (≥10 mutations/megabase) tumours were 1.28 (95% CI 0.72 to 2.28) and 2.56 (95% CI 1.61 to 4.07) for 1–19 and ≥20 pack-years (vs 0 pack-years; P_{trend}<0.001), respectively. In contrast, pack-years smoked were not associated with e-TMB-low tumour incidence (P_{trend}=0.67). A similar differential association was observed for the neoantigen load ($P_{heterogeneity}$=0.017). The differential association by e-TMB appeared consistent in the strata of CpG island methylator phenotype status, *BRAF* mutation or lymphocytic infiltrates.

Conclusions Smoking is more strongly associated with the long-term incidence of colorectal carcinoma harbouring higher mutational and neoantigen loads. Our PCIBM-based evidence supports the immunosuppressive

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Microsatellite instability (MSI)-high status is a clinical biomarker for response to immune checkpoint inhibitors in solid tumours.
⇒ The association between cigarette smoking and colorectal cancer risk is stronger for MSI-high tumours than for non-MSI-high tumours.
⇒ Whether the association of smoking with colorectal cancer incidence is stronger for tumours with higher mutational burdens (or neoantigen loads) independent of the MSI status remains uncertain.

WHAT THIS STUDY ADDS

⇒ Using the prospective cohort incident-tumour biobank method (PCIBM), we found a stronger association of smoking with the incidence of colorectal cancer harbouring higher exome-wide mutation and neoantigen loads.
⇒ Our findings provide unique evidence for the interplay of smoking and tumour somatic mutations (likely influencing antitumour immune response) during tumour development.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ Our PCIBM-based study supports the role of smoking cessation as an immune-enhancing intervention for cancer-free persons and patients with colorectal cancer.
⇒ Further research is warranted to examine the synergistic effect of smoking cessation and the immune checkpoint blockade in patients with cancer.

effect of smoking and the potential of smoking cessation in improving antitumour immunity for cancer prevention and treatment.

INTRODUCTION

Immunotherapy has become a major therapeutic modality in clinical oncology. Microsatellite instability (MSI)-high status currently serves as a reliable clinical biomarker for response to the immune checkpoint blockade in solid tumours.[1–5] The therapeutic responsiveness of MSI-high tumours may be attributable to an increased abundance of immunogenic neopeptides ('neoantigens'), some of which directly stimulate the immune response.[6–9] In addition, tumour mutation burden (TMB, typically measured as the number of nonsynonymous mutations per megabase sequenced) can be used as another tumour biomarker separate from the MSI status.[10 11] Identifying nonsynonymous somatic mutations (and resulting neoantigens) throughout the exome requires data from whole exome, genome or transcriptome sequencing on a pair of tumour and normal specimens. As such, omics sequencing technologies are increasingly utilised in clinical practice. Exome-wide TMB (referred to as 'e-TMB' hereafter) and neoantigens (rather than MSI status) will be clinical biomarkers for response to immunotherapy. For tumour immunogenicity measurements, e-TMB is superior to commonly used TMB based on targeted sequencing assays of selected cancer-associated genes because gain or loss of function driver mutations in those cancer-associated genes have high selection pressures during tumour evolution.[10–12]

Cigarette smoking is a modest risk factor for colorectal carcinomas.[13–16] Although cigarette smoke is a known mutagen linked to the 'smoking mutational signature,'[17 18] this signature has not been observed in colorectal carcinoma. Epidemiological studies have consistently shown a stronger association of smoking with colorectal cancer incidence for MSI-high tumours than for non-MSI-high tumours.[13–16] There appears to be an even stronger association of smoking with MSI-high colorectal carcinoma containing fewer T cells, supporting the immunosuppressive effects of smoking that likely play a pathogenic role in MSI-high colorectal tumours.[19] Considering that neoantigens (rather than MSI-high status per se) directly stimulate antitumour immune response, we hypothesised that the association of smoking with long-term colorectal cancer incidence might be stronger for tumours with higher levels of e-TMB and neoantigen loads.

Using two large prospective cohort studies in the USA with data on long-term smoking habit and whole-exome sequencing (WES) of colorectal cancer and matched normal tissue, we examined longitudinally updated pack-years smoked and the long-term incidence of colorectal carcinomas subclassified by e-TMB or neoantigen loads. We applied the prospective cohort incident-tumour biobank method (PCIBM)[20 21] to decades-long prospective collection of comprehensive lifestyle data and tumour genomic profiling. The current study represents the first prospective investigation of longitudinally updated smoking habit in relation to the long-term incidence of colorectal carcinomas classified by somatic genomic profiles based on WES.

MATERIALS AND METHODS

Study population

We used the PCIBM[20 21] on data from two ongoing prospective cohort studies in the USA, the Nurses' Health Study (NHS, 121 700 women aged 30–55 years followed since 1976) and the Health Professionals Follow-up Study (HPFS, 51 529 men aged 40–75 years followed since 1986) (table 1 and figure 1).[22 23] Using mailed biennial questionnaires, participants have reported lifestyle factors including smoking behaviour and newly diagnosed diseases. The response rate has exceeded 90% for each follow-up questionnaire cycle in both cohorts. At the baseline (1980 for the NHS and 1986 for the HPFS), we excluded participants who did not return the initial food frequency questionnaire, left a large number of items blank (>10 of 61 items for the NHS and >70 of 131 items for the HPFS), reported unreasonable total calorie intake (<600 or >3500 calories/day for women, and <800 or >4200 calories/day for men), or reported a history of inflammatory bowel disease. We additionally excluded participants with a history of cancer except for non-melanoma skin cancer to rule out the possibility of metastatic tumours to the colorectum and that of biases derived from lifestyle alterations due to cancer diagnosis. Participants were followed until death or the end of follow-up (1 June 2012 for the NHS; and 31 January 2012 for the HPFS), whichever came first.

Assessment of smoking behaviour

The details of smoking behaviour were assessed as reported previously.[19 24] In 1976 (the NHS) and 1986 (the HPFS), participants reported the age when they began smoking (and ceased smoking, if applicable), as well as the average daily consumption of cigarettes. Participants have updated current smoking status and daily cigarette consumption every 2 years. We calculated cumulative pack-years smoked (average daily consumption of cigarette packs multiplied by the number of years smoked) at the baseline and every 2 years thereafter.

Acquisition of colorectal cancer cases

In both cohorts, colorectal carcinoma cases were identified based on biennial questionnaires. For non-respondents, colorectal cancer cases and deaths were ascertained through family members, US post office authorities and/or the National Death Index. Study physicians, blinded to exposure data, reviewed medical records of identified colorectal cancer cases to confirm the diagnosis and record tumour characteristics (eg, anatomical location and disease stage). We included both colon and rectal carcinomas based on the colorectal continuum model.[25] We collected

Table 1 Age-standardised characteristics of participants according to cumulative pack-years smoked in the Nurses' Health Study (NHS, 1980–2012) and the Health Professionals Follow-up Study (HPFS, 1986–2012)

	Women (NHS)			Men (HPFS)		
	Cumulative pack-years smoked			Cumulative pack-years smoked		
Characteristic*	0	1–19	≥20	0	1–19	≥20
Person-years	1 125 146	705 774	661 296	483 866	231 160	266 200
Age, years	60.9 (11.5)	59.4 (11.5)	61.6 (10.7)	63.2 (11.3)	63.3 (11.1)	66.6 (10.6)
Family history of colorectal cancer	13%	14%	13%	13%	12%	12%
History of diabetes	7.4%	6.8%	7.8%	6.8%	7.3%	9.3%
Body mass index, kg/m²	25.5 (4.7)	25.2 (4.6)	25.1 (4.5)	25.6 (3.4)	25.7 (3.2)	26.3 (3.6)
Postmenopause	76%	76%	82%	–	–	–
Menopausal hormone therapy	28%	29%	24%	–	–	–
History of colonoscopy/sigmoidoscopy	40%	43%	36%	54%	56%	50%
Regular use of multivitamins	53%	54%	48%	44%	45%	43%
Regular use of aspirin	39%	40%	41%	46%	49%	49%
Regular use of other NSAIDs	17%	19%	18%	15%	17%	16%
Physical activity, METS-hours/week	16.5 (16.8)	18.0 (18.8)	15.1 (16.3)	26.9 (23.6)	27.0 (22.6)	22.3 (21.1)
Total calorie intake, kcal/day	1702 (443)	1677 (432)	1645 (439)	1983 (554)	1965 (550)	1980 (560)
Alcohol intake, g/day	3.8 (6.9)	6.9 (8.9)	8.7 (11.8)	8.0 (11.1)	12.5 (13.8)	15.1 (16.9)
Red and processed meat intake, servings/week	6.6 (3.7)	6.3 (3.5)	6.8 (3.7)	6.1 (4.3)	6.1 (4.2)	7.2 (4.8)
Total calcium intake, mg/day	939 (357)	953 (352)	887 (350)	957 (375)	932 (367)	897 (374)
Total folate intake, µg/day	429 (212)	439 (212)	398 (204)	552 (253)	560 (257)	511 (250)
Alternate Healthy Eating Index 2010†	46.1 (9.6)	47.5 (9.6)	45.3 (9.6)	48.6 (10.1)	49.0 (9.9)	46.6 (10.1)

*All variables other than age were standardised to age distribution of each cohort. Mean (SD) was presented for continuous variables.
†Without alcohol intake.
METS, Metabolic Equivalent Task Score; NSAID, non-steroidal anti-inflammatory drug.

formalin-fixed paraffin-embedded (FFPE) tissue blocks of surgically resected colorectal tumours from hospitals throughout the USA, and the study pathologist (SO) confirmed a pathological diagnosis of colorectal carcinoma. During the follow-up of the participants, we documented 3053 incident colorectal cancer cases, including 752 cases with available WES data. There were no substantial differences in clinical data between cases with and without WES data (online supplemental table S1).

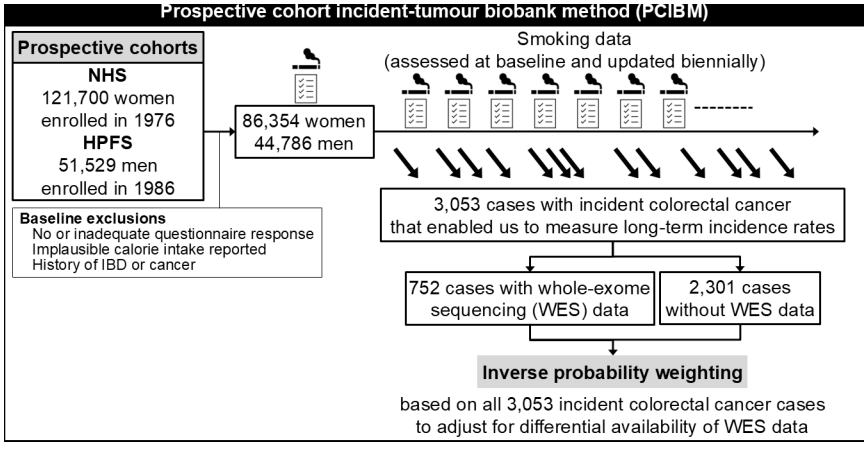

Figure 1 Flow diagram of the study population in the Nurses' Health Study (NHS) and the Health Professionals Follow-up Study (HPFS). In the current study based on the PCIBM, we included 86 354 women and 44 786 men who were followed for decades and examined the incidence of colorectal carcinomas subclassified by exome-wide tumour mutational burden or neoantigen loads. To reduce selection bias due to tissue availability, we applied the inverse probability weighting for 752 cases with available WES data using the data from 3053 cases. IBD, inflammatory bowel disease.

Analyses of colorectal cancer tissue

The study pathologist (SO), blinded to other data, reviewed H&E-stained tissue sections and recorded pathological features including four patterns of lymphocytic reaction (tumour-infiltrating lymphocytes, intratumoural periglandular reaction, peritumoural lymphocytic reaction and Crohn's-like lymphoid reaction).[26 27] Tumour status of MSI, CIMP and *BRAF* mutation was assessed as previously described.[25 28 29] Tumour MSI status was assessed using PCR of 10 microsatellite markers (D2S123, D5S346, D17S250, BAT25, BAT26, BAT40, D18S55, D18S56, D18S67 and D18S487), and MSI-high was defined as the presence of instability in ≥30% of the markers.[25]

WES and downstream analyses

The study pathologist (SO) marked tumour areas in guide H&E-stained slides. Using the guide H&E slides to ensure high tumour cellularity, DNA was extracted from tumour areas in sections of archival FFPE blocks.[30] Matched normal DNA was obtained from normal colon tissue that was grossly away from the tumour. As previously described,[31] we performed WES on DNA from tumour and matched normal tissue pairs with the mean target coverage of 85× and the mean of 49 million paired-end reads across all samples. To remove artefacts resulting from hydrolytic deamination of cytosine to uracil in FFPE samples, we filtered out C to T transition mutations as possible FFPE-specific artefacts using the tool described elsewhere.[32] To further filter out spurious single-nucleotide variant calls, we used BWA (Burrows-Wheeler Aligner)-MEM (http://bio-bwa.sourceforge.net/) to realign sequenced reads associated with the mutations to a set of sequences derived from the human reference assembly. e-TMB was defined as the number of non-synonymous mutations per megabase covered in the whole exome. We also calculated TMB based on the panel of selected cancer-associated genes (called 'targeted TMB'), that is, 447 genes used in clinical practice at the Brigham and Women's Hospital (listed in online supplemental table S2). The neoantigen load (ie, the number of mutated proteins that likely give rise to immunogenic peptides) was calculated by counting peptides that were predicted to bind to HLA molecules with high affinity (the rank <0.5%), as previously described.[33] Using NetMHCpan (V.4.1),[33] we predicted the binding affinities of 9-mer and 10-mer mutant peptides found in tumours to the corresponding *HLA* alleles inferred by the POLYSOLVER algorithm.[34] e-TMB was categorised as high and low at the cut-off point of 10 mutations per megabase that was adopted for the US Food and Drug Administration approval of pembrolizumab for TMB-high tumours and has been commonly used.[35 36] Given that TMB-high cases represented 13% of all cases with available WES data, the neoantigen load was categorised as high (≥326 per exome, the top quartile) and low (<326 per exome, the other quartiles). We examined the single-base substitution (SBS) signatures,[17] including SBS11 (alkylating signature) associated with red meat consumption in individuals developing colorectal cancer.[31] SBS4 (smoking signature) characterised by C–A nucleotide transversions was specifically examined.

Statistical analysis

All statistical analyses were performed using SAS software (V.9.4, SAS Institute), and all p values were two-sided. We used the two-sided α level of 0.005, as recommended by expert statisticians.[37] Our primary hypothesis testing was an assessment of the heterogeneity between associations of cumulative pack-years smoked (continuous with a ceiling at 50 pack-years) with the incidence of colorectal cancer subclassified by e-TMB or neoantigen loads (high vs low).

We used the Cox proportional hazards regression model to estimate the HR for colorectal cancer incidence. To assess differential associations of smoking variables with the incidence of colorectal cancer subclassified by e-TMB or neoantigen loads, we used the duplication-method Cox regression model for competing risks.[38] Using the likelihood ratio test (1 df), we examined a heterogeneity trend across tumour subtypes in a statistical trend of the association across smoking exposure levels.[39] The multivariable Cox regression model included the covariates described in table 2. We treated all exposure variables as time-dependent to account for changes over time. To reduce intraindividual variation and consider long-term influences, we used the cumulative average for relevant variables, which was the mean of all available data prior to each questionnaire cycle. The Cox regression models were stratified by sex (for pooled analyses), age and calendar year of the questionnaire cycle. Colorectal cancer cases without WES data were treated as censored at the time of cancer diagnosis.

To adjust for selection bias caused by the availability of WES data, we used the inverse probability weighting (IPW) method (figure 1).[40] Using covariate data on the 3053 incident colorectal cancer cases, we constructed the multivariable logistic regression model to calculate the cohort-specific probability of WES data availability in each patient. In the IPW-adjusted Cox regression model, each colorectal cancer patient with available WES data was weighted by the inverse of the probability. For example, when a patient with colorectal cancer with available WES data was estimated (based on the patient's statuses of covariates) to have the data with a probability of 0.8, this patient was weighted by the inverse probability (ie, 1/0.8=1.25). Through this statistical approach, a bias due to the differential availability of WES data according to the covariate statuses was mitigated, increasing the representativeness of our sample of colorectal cancer cases and enhancing the generalisability of our results. Weights greater than the 95th percentile were truncated and set to the 95th percentile to reduce outlier effects. We confirmed that results with and without weight truncation did not differ substantially (data not shown).

We conducted tests of heterogeneity between the two cohorts using the *Q* statistic and observed no statistically

Table 2 Cumulative pack-years smoked and colorectal cancer incidence, overall and by exome-wide tumour mutational burden or neoantigen loads

	Cumulative pack-years smoked				
	0	1–19	≥20	P_{trend}*	$P_{heterogeneity}$†
Person-years	1 609 012	936 934	927 496		
All colorectal cancer (n=752)					
n	316	189	247		
Age-adjusted HR (95% CI)‡	1 (referent)	1.12 (0.94 to 1.35)	1.20 (1.02 to 1.43)	0.012	–
Multivariable HR (95% CI)‡§	1 (referent)	1.15 (0.95 to 1.38)	1.16 (0.97 to 1.37)	0.080	–
Exome-wide tumour mutational burden¶					<0.001
Low (n=654)					
n	286	169	199		
Age-adjusted HR (95% CI)‡	1 (referent)	1.11 (0.92 to 1.34)	1.08 (0.90 to 1.29)	0.36	
Multivariable HR (95% CI)‡§	1 (referent)	1.14 (0.94 to 1.38)	1.05 (0.87 to 1.27)	0.67	
High (n=98)					
n	30	20	48		
Age-adjusted HR (95% CI)‡	1 (referent)	1.24 (0.69 to 2.22)	2.61 (1.64 to 4.15)	<0.001	
Multivariable HR (95% CI)‡§	1 (referent)	1.28 (0.72 to 2.28)	2.56 (1.61 to 4.07)	<0.001	
Neoantigen loads¶					0.017
Low (n=564)					
n	242	154	168		
Age-adjusted HR (95% CI)‡	1 (referent)	1.19 (0.97 to 1.45)	1.06 (0.87 to 1.30)	0.34	
Multivariable HR (95% CI)‡§	1 (referent)	1.22 (0.99 to 1.49)	1.03 (0.84 to 1.27)	0.62	
High (n=188)					
n	74	35	79		
Age-adjusted HR (95% CI)‡	1 (referent)	0.91 (0.60 to 1.36)	1.76 (1.27 to 2.42)	<0.001	
Multivariable HR (95% CI)‡§	1 (referent)	0.95 (0.63 to 1.42)	1.73 (1.25 to 2.39)	0.002	

*P_{trend} was calculated using a linear trend test and cumulative pack-years smoked (continuous with a ceiling at 50 pack-years).
†$P_{heterogeneity}$ was calculated using the likelihood ratio test (1 df) for the heterogeneity of binary subtype-specific associations of cumulative pack-years smoked (continuous with a ceiling at 50 pack-years) in multivariable models.
‡Inverse probability weighting was applied to reduce a potential selection bias due to the differential availability of whole-exome sequencing data (see 'Statistical analysis' subsection for details).
§The multivariable Cox regression model included family history of colorectal cancer (present vs absent), body mass index (continuous with a ceiling at 35 kg/m^2), history of colonoscopy/sigmoidoscopy (present vs absent), use of aspirin or other non-steroidal anti-inflammatory drugs (regular use vs non-use), physical activity (continuous with a ceiling at 50 metabolic equivalent task score-hours/week), alcohol intake (continuous with a ceiling at 30 g/day), red and processed meat intake (continuous with a ceiling at 14 servings/week) and total folate intake (continuous with a ceiling at 1000 µg/day). For women, we additionally included menopause status/menopausal hormone therapy (premenopause vs postmenopause with never, past or current use of menopausal hormone therapy).
¶e-TMB was categorised into high (≥10 per megabase) and low (<10 per megabase). Based on all colorectal cancer cases with available whole-exome sequencing data, neoantigen loads were categorised into high (≥326 per exome, the top quartile) and low (<326 per exome, the other quartiles).
e-TMB, exome-wide tumour mutational burden.

significant heterogeneity ($P_{heterogeneity}$>0.009) in the associations between cumulative pack-years smoked and the incidence of colorectal cancer subclassified by e-TMB or neoantigen loads. We, therefore, combined the cohorts with the adjustment for cohort (ie, sex) for further analyses to increase statistical power.

Patient and public involvement

Patients and/or the public were not involved in the design, conduct, reporting or dissemination plans of this research.

RESULTS

Table 1 summarises age-standardised characteristics of the cohort participants (the histograms of cumulative pack-years smoked are presented in online supplemental figure S1). During 3 473 441 person-years follow-up of 131 140 participants, we documented 3053 colorectal cancer cases including 752 cases with available WES data (online supplemental figure S2), which yielded e-TMB (non-synonymous mutation count per megabase; median, 1.6; IQR, 1.0–3.5 and total range, 0–152.0) and

neoantigen loads (median, 202; IQR, 137–326 and total range, 0–11,110). Tumour MSI status, e-TMB and neoantigen loads correlated with each other (online supplemental figure S3). Cumulative pack-years smoked were not associated with the smoking signature (online supplemental figure S4). The levels of e-TMB and neoantigen loads were not correlated with the smoking signature (online supplemental figure S5). Cumulative pack-years smoked were associated with the incidence of colorectal cancer using all of the incident cases, and this association appeared persistent regardless of WES data availability (online supplemental table S3). For further incidence analyses (except for a sensitivity analysis), we used the IPW method and the 3053 cases to adjust for selection bias due to tissue data availability.

In our primary hypothesis testing, the association of cumulative pack-years smoked with colorectal cancer incidence differed by e-TMB ($P_{heterogeneity}$<0.001, table 2 and figure 2). Compared with never smokers, multivariable-adjusted HRs for TMB-high colorectal cancer in individuals who smoked 1–19 and ≥20 pack-years were 1.28 (95% CI 0.72 to 2.28) and 2.56 (95% CI, 1.61 to 4.07), respectively (P_{trend}<0.001). In contrast, cumulative pack-years smoked were not significantly associated with the incidence of colorectal cancer containing lower levels of e-TMB (P_{trend}=0.67). A similar differential association was observed for neoantigen loads ($P_{heterogeneity}$=0.017, table 2). Our findings in each cohort are presented in online supplemental table S4. Our sensitivity analyses without IPW adjustment yielded similar results to those with IPW adjustment (online supplemental table S5). As targeted sequencing of selected cancer-associated genes has become common in clinical practice, TMB measures calculated from such targeted sequencing analyses are increasingly used. Thus, we calculated 'targeted TMB' based on a clinical panel of cancer-associated genes used in the Brigham and Women's Hospital, which was only moderately correlated with e-TMB (correlation coefficient, 0.62; online supplemental figure S6). We conducted analyses using 'targeted TMB' and found similar but attenuated results compared with those using e-TMB (online supplemental table S6).

In secondary analyses, we examined smoking status (current vs past vs never) and duration of smoking cessation in relation to the incidence of colorectal carcinomas subclassified by e-TMB or neoantigen loads. Compared with never smoking, former and current smoking was associated with a higher incidence of colorectal cancer harbouring high levels of e-TMB but not with colorectal cancer harbouring lower levels of e-TMB ($P_{heterogeneity}$=0.003, online supplemental table S7). Similarly, an association of the duration of smoking cessation with the incidence of colorectal carcinomas was stronger for carcinomas harbouring high levels of e-TMB ($P_{heterogeneity}$=0.001, online supplemental table S8).

We conducted secondary subgroup analyses. In an analysis stratified by tumour MSI status, we observed a stronger association of cumulative pack-years smoked with the incidence of tumours containing higher levels of e-TMB or neoantigen loads in the stratum of MSI-high tumours (table 3), though the differential associations did not reach statistical significance. Our previous analyses suggest that smoking status may be differentially associated with the incidence of colorectal cancer by tumour status of CIMP, *BRAF* mutation and T lymphocyte infiltrates.[14 19] Therefore, we conducted additional analyses stratified by tumour status of CIMP, *BRAF* mutation or lymphocytic reaction, which yielded similar differential associations, though statistical power was limited in the respective strata (online supplemental table S9).

DISCUSSION

Because colorectal carcinoma is a group of neoplasms that evolve through heterogeneous sets of genetic and epigenetic alterations influenced by lifestyle and environmental factors,[41] the molecular pathological epidemiology (MPE) approach is useful to gain insights into the interplay of lifestyle exposures and tumour molecular alterations during the tumourigenic process.[42–44] Using the PCIBM on data on long-term smoking habit and tumour WES,[20 21] our study has shown that longitudinally updated pack-years smoked are associated with higher long-term incidence of colorectal cancer harbouring high levels of e-TMB or neoantigen loads, but not colorectal cancer containing low levels of these parameters. Such a differential association appeared to persist in MSI-high colorectal cancer. The MSI-high status has been shown to influence immune response through non-synonymous mutations and neoantigen production.[6 45 46] Our findings suggest that tobacco compounds may make an immunosuppressive microenvironment that can favour the growth of colorectal tumours with a high burden of nonsynonymous mutations and neoantigens, providing further evidence for the suppressive effect of smoking on antitumour immunity.

The immune checkpoint inhibitors that target the CD274 (PDCD1 ligand 1, PD-L1)-PDCD1 (programmed cell death 1, PD-1) coinhibitory pathway have shown great promise in treating tumours, but their effectiveness has been confined to a limited subset of tumours including MSI-high tumours.[1–5] Certain neoantigens (ie, cancer-specific antigens resulting from somatic mutations) are considered to be presented by major histocompatibility complex molecules and recognised as non-self-epitopes by T lymphocytes.[7 8] Hence, compared with the MSI-high status, TMB and neoantigen loads are considered to be better predictors for tumour immunogenicity and responsiveness to the immune checkpoint blockade. Emerging evidence supports the predictive ability of TMB for clinical benefits from the immune checkpoint inhibitors in non-MSI-high tumours.[47–49] Given the increasing availability of tumour omics profiling in clinical practice, e-TMB and neoantigen loads will likely replace tumour MSI status as a biomarker for the effectiveness of the immune checkpoint blockade.

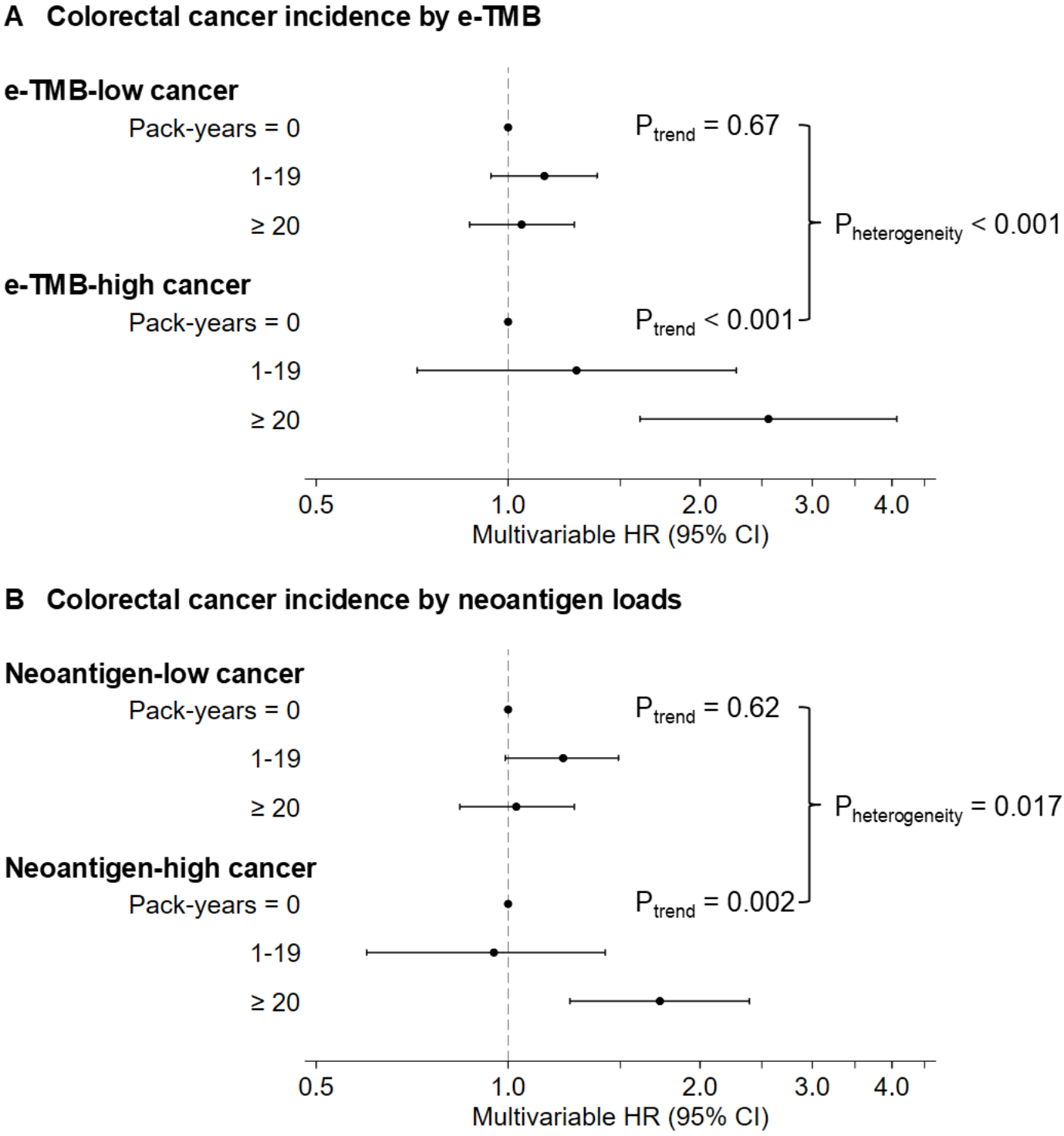

Figure 2 Forest plot of multivariable HRs for the incidence of colorectal cancer classified by exome-wide tumour mutational burden (A) or neoantigen loads (B) according to cumulative pack-years smoked. The dot indicates a stratum-specific multivariable HR, and the horizontal bar indicates 95% CI. The HRs were adjusted for the same set of covariates as table 2, and inverse probability weighting was applied to reduce a potential selection bias due to the differential availability of whole-exome sequencing data (see 'Statistical analysis' subsection for details). P_{trend} was calculated using a linear trend test and cumulative pack-years smoked (continuous with a ceiling at 50 pack-years). $P_{heterogeneity}$ was calculated using the likelihood ratio test (1 df) for the ordinal trend heterogeneity of quartile subtype-specific associations of cumulative pack-years smoked (continuous with a ceiling at 50 pack-years). e-TMB, exome-wide tumour mutational burden.

Cigarette smoke contains thousands of DNA mutagens, which are considered to cause a distinct mutational pattern (referred to as the 'smoking signature') characterised by a high frequency of C–A nucleotide transversions.[17 18] However, the smoking mutational signature commonly observed in lung carcinoma has not been well

Table 3 Cumulative pack-years smoked and colorectal cancer incidence by exome-wide tumour mutational burden or neoantigen loads in the strata of tumour microsatellite instability (MSI) status

	Cumulative pack-years smoked				
	0	1–19	≥20	P_{trend}*	$P_{heterogeneity}$†
MSI-high					
Exome-wide tumour mutational burden‡					0.14
Low (n=61)					
n	20	15	26		
Age-adjusted HR (95% CI)§	1 (referent)	1.35 (0.69 to 2.65)	2.00 (1.11 to 3.62)	0.053	
Multivariable HR (95% CI)§¶	1 (referent)	1.35 (0.68 to 2.65)	2.01 (1.11 to 3.66)	0.045	
High (n=60)					
n	18	10	32		
Age-adjusted HR (95% CI)§	1 (referent)	1.10 (0.50 to 2.44)	2.94 (1.63 to 5.29)	< 0.001	
Multivariable HR (95% CI)§¶	1 (referent)	1.11 (0.50 to 2.44)	2.99 (1.63 to 5.49)	< 0.001	
Neoantigen loads‡					0.22
Low (n=60)					
n	22	10	28		
Age-adjusted HR (95% CI)§	1 (referent)	0.83 (0.39 to 1.78)	1.96 (1.11 to 3.45)	0.036	
Multivariable HR (95% CI)§¶	1 (referent)	0.83 (0.39 to 1.77)	1.97 (1.10 to 3.51)	0.034	
High (n=61)					
n	16	15	30		
Age-adjusted HR (95% CI)§	1 (referent)	1.76 (0.86 to 3.61)	3.08 (1.67 to 5.71)	< 0.001	
Multivariable HR (95% CI)§¶	1 (referent)	1.77 (0.87 to 3.61)	3.15 (1.68 to 5.90)	< 0.001	
Non-MSI-high					
Exome-wide tumour mutational burden‡					0.73
Low (n=292)					
n	118	96	78		
Age-adjusted HR (95% CI)§	1 (referent)	1.51 (1.15 to 1.97)	1.04 (0.78 to 1.39)	0.47	
Multivariable HR (95% CI)§¶	1 (referent)	1.56 (1.19 to 2.04)	1.00 (0.74 to 1.34)	0.83	
High (n=292)					
n	137	60	95		
Age-adjusted HR (95% CI)§	1 (referent)	0.83 (0.61 to 1.13)	1.05 (0.81 to 1.37)	0.85	
Multivariable HR (95% CI)§¶	1 (referent)	0.86 (0.63 to 1.16)	1.01 (0.77 to 1.32)	0.80	
Neoantigen loads‡					0.18
Low (n=292)					
n	117	88	87		
Age-adjusted HR (95% CI)§	1 (referent)	1.43 (1.08 to 1.88)	1.15 (0.87 to 1.52)	0.15	
Multivariable HR (95% CI)§¶	1 (referent)	1.46 (1.11 to 1.93)	1.10 (0.82 to 1.46)	0.36	
High (n=292)					
n	138	68	86		
Age-adjusted HR (95% CI)§	1 (referent)	0.93 (0.69 to 1.24)	0.97 (0.74 to 1.27)	0.62	
Multivariable HR (95% CI)§¶	1 (referent)	0.96 (0.72 to 1.29)	0.93 (0.71 to 1.23)	0.35	

*P_{trend} was calculated using a linear trend test and cumulative pack-years smoked (continuous with a ceiling at 50 pack-years).
†$P_{heterogeneity}$ was calculated using the likelihood ratio test (1 df) for the heterogeneity of binary subtype-specific associations of cumulative pack-years smoked (continuous with a ceiling at 50 pack-years) in multivariable models.
‡The number of MSI-high tumours containing low levels of e-TMB (or neoantigen loads) was quite small. Therefore, e-TMB was categorised into high and low based on stratum-specific median cut-off points (13 and 1.3 per megabase for MSI-high and non-MSI-high tumours, respectively). Similarly, neoantigen loads were categorised into high and low based on stratum-specific median cut-off points (2956 and 177 per exome for MSI-high and non-MSI-high tumours, respectively).
§Inverse probability weighting was applied to reduce a potential selection bias due to the differential availability of whole-exome sequencing data (see 'Statistical analysis' subsection for details).
¶The multivariable Cox regression model was adjusted for the same set of covariates as table 2.
e-TMB, exome-wide tumour mutational burden.;

described in colorectal carcinoma. Tobacco compounds may promote carcinogenesis in various organ systems, and their mutagenic and immunosuppressive effects have been proposed as mechanisms of smoking-induced carcinogenesis.[50–53] Nicotine, a major component of cigarette smoke, has been shown to impair the functions of dendritic cells and NK cells, which may promote the immune evasion of tumour cells.[51] A previous study has shown a stronger association of smoking with colorectal cancer incidence for tumours with lower-level T cell infiltrates.[19] Our study suggests that the immunosuppressive effect of smoking may contribute to the proliferation and survival of highly immunogenic tumour cells with abundant somatic mutations. Our analyses suggest that cumulative pack-years smoked or tumour mutational and neoantigen loads were not strongly correlated with the smoking signature in colorectal cancer. These findings further suggest that the carcinogenic effect of cigarette smoke is independent of its role in inducing mutations in the colorectal epithelium, which contrasts with the lung epithelium where smoking exerts its mutagenic effects more directly.[17 18] Future research should examine whether prediagnosis smoking status is associated with the effectiveness of immunotherapy for colorectal cancer and whether the cessation of smoking may improve the therapeutic efficacy. Given the stronger association between cumulative pack-years smoked (or smoking cessation (inverse association)) and the incidence of colorectal cancer having higher immunogenicity, immune checkpoint inhibitors with preceding or in place of cytotoxic chemotherapy may be tested for efficacy in smokers diagnosed with TMB-high colorectal cancer. In addition, smoking cessation may be tested to determine whether it can augment the effect of immune checkpoint inhibitors for TMB-high colorectal cancer.

The current study conducted integrative MPE analyses to assess risk factor exposures and tumour molecular features.[42 43 54 55] This MPE approach has been used to examine smoking in relation to cancer incidence/risk by tumour subtypes according to molecular pathology,[13–16 56] faecal microbiome[57] and immune cell infiltrates,[19 24 58] to shed light on the carcinogenic effects of tobacco smoke. Furthermore, the current study took advantage of the PCIBM,[20 21] which has allowed for assessment of long-term exposures/cancer incidence by tumour subtypes.[13–16 19 24 56 58]

Obtaining data on exome-scale TMB and neoantigen loads requires high-throughput sequencing technologies and computational analysis platforms such that no prior prospective study has examined the association of epidemiological factors with the incidence of colorectal cancer subclassified by e-TMB or neoantigen loads.[10 11] In clinical practice, targeted sequencing assays for selected cancer-associated gene panels are increasingly used, and the term ‘TMB’ is commonly used for a mutational frequency in the selected genes. However, pathogenic mutations in those selected cancer-associated genes are subject to substantial selection pressure, which causes imprecision when estimating the e-TMB, neoantigen load and tumour immunogenicity.[10 11] In addition, gene panels in targeted sequencing assays documented in the literature differ from study to study, which makes cross-comparisons challenging. To avoid these problems, we used WES to capture exome-wide mutational profile and better estimate the tumour immunogenicity. Because of the increasing availability and reducing costs of omics assays, personalised treatment of cancer patients based on comprehensive exomic mutational profiling will likely be implemented in the future. In the I-PREDICT studies, patients with treatment-naïve or refractory malignancy were treated with agents matched for identified molecular alterations, and targeting a larger fraction of the alterations resulted in better survival outcomes.[59 60] In parallel with the trend of precision oncology, our study highlights the potential of the exome-wide assessment of tumour immunogenicity in epidemiological research.

The current study has limitations. First, there is the possibility of unmeasured and/or residual confounding. Nonetheless, we adjusted for many established risk factors of colorectal cancer in our multivariable models, and this adjustment did not alter our findings materially from those of the univariable analyses. It is also noted that randomising smoking exposure in a trial study to eliminate confounding would be unethical and thus impossible. Second, the neoantigen loads were estimated using in silico methods that depended on HLA class I predictions. Using the method, a previous study found that the neoantigen load from WES data robustly correlated with lymphocytic reaction levels in colorectal cancer.[6] Nonetheless, the data on neoantigen load had certain measurement errors and should be replicated in independent studies. Moreover, we could examine e-TMB as another measure of tumour immunogenicity, which yielded similar data to those using the neoantigen loads. Third, WES data were not available for many colorectal cancer cases within the cohorts, which might have caused selection bias. However, we used all of the 3053 incident colorectal cancers and the IPW method to adjust for the selection bias,[40] and analyses with and without the IPW adjustment yielded similar results. Fourth, most study participants were Caucasian health professionals, and therefore, our findings need to be validated in independent populations.

The current study has notable strengths. First, our prospective cohort design enabled us to obtain longitudinally updated data on smoking habits and potential confounders, while eliminating differential recall bias between cancer cases and cancer-free individuals. Second, the prospective study design also enabled us to use the 3053 incident colorectal cancers to adjust for selection bias due to the availability of the WES data. Third, in contrast to commonly used targeted sequencing assays, WES analyses could yield exome-scale TMB and neoantigen loads. Fourth, integrated MPE analyses using the PCIBM on the prospective cohort studies with the WES data in incident tumours could assess not only the

longitudinal effect of smoking on the long-term incidence of WES-based tumour subtypes but also statistical heterogeneity between the subtypes. Fifth, our database also enabled us to evaluate results after controlling for other key tumour characteristics such as MSI status and T cell infiltrates. Sixth, our incident colorectal cancer cases represented a collection of patients who visited hundreds of different hospitals throughout the USA, which provided higher generalisability compared with studies based on only a limited number of hospitals.

CONCLUSIONS

The current study demonstrated a stronger association of smoking with colorectal cancer incidence for tumours containing higher numbers of exome-wide somatic mutations. Smoking may contribute to the development of colorectal tumours, especially those with high frequencies of somatic mutations, possibly through its effect on the tumour immune microenvironment. Future studies should examine whether cessation of smoking may stimulate antitumour immune response and improve response to immune checkpoint inhibitors. As omics analysis platforms are increasingly available and cost-efficient for routine tumour pathological testing, the current study will inform the effort of implementing tumour exome sequencing analyses in epidemiological research and clinical practice.

Author affiliations
[1]Program in MPE Molecular Pathological Epidemiology, Department of Pathology, Brigham and Women's Hospital, Boston, Massachusetts, USA
[2]Department of Gastroenterology, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan
[3]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA
[4]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts, USA
[5]Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA
[6]Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA
[7]Yale University School of Nursing, Orange, Connecticut, USA
[8]Department of Epidemiology and Biostatistics, and the Ministry of Education Key Lab of Environment and Health, Huazhong University of Science and Technology, Hubei, China
[9]Division of Public Health Sciences, Department of Surgery, Washington University School of Medicine, St. Louis, Missouri, USA
[10]Clinical and Translational Epidemiology Unit, Massachusetts General Hospital, Boston, Massachusetts, USA
[11]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA
[12]Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts, USA
[13]Department of Hematopoietic Biology and Malignancy, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA
[14]Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts, USA
[15]Genentech Inc, South San Francisco, California, USA
[16]Division of Gastroenterology, Massachusetts General Hospital, Boston, Massachusetts, USA
[17]Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA
[18]Cancer Immunology and Cancer Epidemiology Programs, Dana-Farber/Harvard Cancer Center, Boston, Massachusetts, USA
[19]Institute of Science Tokyo, Tokyo, Japan

Acknowledgements The authors would like to acknowledge the contribution to this study from central cancer registries supported through the Centers for Disease Control and Prevention's National Program of Cancer Registries (NPCR) and/or the National Cancer Institute's Surveillance, Epidemiology and End Results (SEER) Programme. Central registries may also be supported by state agencies, universities and cancer centres. Participating central cancer registries include the following: Alabama, Alaska, Arizona, Arkansas, California, Colorado, Connecticut, Delaware, Florida, Georgia, Hawaii, Idaho, Indiana, Iowa, Kentucky, Louisiana, Massachusetts, Maine, Maryland, Michigan, Mississippi, Montana, Nebraska, Nevada, New Hampshire, New Jersey, New Mexico, New York, North Carolina, North Dakota, Ohio, Oklahoma, Oregon, Pennsylvania, Puerto Rico, Rhode Island, Seattle SEER Registry, South Carolina, Tennessee, Texas, Utah, Virginia, West Virginia and Wyoming. The authors assume full responsibility for analyses and interpretation of these data.

Contributors TH, TU, CG, SU, MG and SO developed the main concept, designed the study and participated in the literature search. XZ, CSF, ATC, MG and SO produced financial support. TH, TU, CG, SU, XZ, KH, YT, NA, MCL, KM, NN, MH, SM, KK, YM, LL, YC, DN, MW, RN, SAS, CJW, LAG, JAM, ELG, JAN, CSF, ATC, MS, MG and SO were responsible for collection of tumour tissue, and acquisition of epidemiologic, clinical and tumour tissue data, including genetic, histopathological, and immunohistochemical characteristics. TH, TU, CG, SU, DN, MW, ATC, MS, MG and SO analysed and interpreted the data. TH, TU, CG, SU, XZ, MS, MG and SO drafted the manuscript. RN, CSF, ATC, MG and SO contributed to editing and critical revision for important intellectual contents. TH, TU, CG, SU, MG and SO had full access to all the data in this study and take full responsibility for the integrity of the data and the accuracy of the data analysis. MG and SO are the guarantors of the study. All authors read and approved the final version of the report. The corresponding authors attest that all listed authors meet the authorship criteria and that no others meeting the criteria have been omitted.

Funding This work was supported by the US National Institutes of Health (NIH) grants (P01 CA87969 to MJ Stampfer, UM1 CA186107 to MJ Stampfer, P01 CA55075 to WC Willett, UM1 CA167552 to WC Willett, U01 CA167552 to WC Willett and LA Mucci, R50 CA274122 to TU, K07 CA188126 to XZ, P50 CA127003 to CSF, R01 CA118553 to CSF, R01 CA169141 to CSF, R01 CA137178 to ATC, K24 DK098311 to ATC, R35 CA197735 to SO, R01 CA151993 to SO and R01 CA248857 to SO); by Cancer Research UK Grand Challenge Award (OPTIMISTICC, C10674/A27140 to MG and SO); by American Cancer Society Clinical Research Professor Award (CRP-24-1185864-01-PROF to SO); by the Stand Up to Cancer (SU2C) Colorectal Cancer Dream Team Translational Research Grant (SU2C-AACR-DT22-17 to CSF and MG), administered by the American Association for Cancer Research, a scientific partner of SU2C; by Nodal Award (2016-02) from the Dana-Farber Harvard Cancer Center (to SO); 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. TU was supported by a fellowship grant from Yasuda Medical Foundation, Brigham and Women's Hospital Faculty Career Development Award, and an Investigator Initiated Grant from the American Institute for Cancer Research (AICR). SU and KH were supported by fellowship grants from the Uehara Memorial Foundation. TU and KK were supported by Overseas Research Fellowship grants from Japan Society for the Promotion of Science (201960541 to TU and JP2017-775 to KK). KH was supported by a fellowship grant from the Mitsukoshi Health and Welfare Foundation. LL was supported by a scholarship grant from the Chinese Scholarship Council and a fellowship grant from Huazhong University of Science and Technology. ATC is a Stuart and Suzanne Steele MGH Research Scholar. MS was supported by the 2017 AACR-AstraZeneca Fellowship in Immuno-oncology Research (17-40-12-SONG). MG was supported by a Conquer Cancer Foundation of ASCO Career Development Award.

Disclaimer The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The funding source had no role in the design and conduct of the study; collection, management, analysis and interpretation of the data; preparation, review, and approval of the manuscript; and decision to submit the manuscript for publication.

Competing interests All authors have completed the ICMJE uniform disclosure form at www.icmje.org/disclosure-of-interest/ and declare the following competing interests. RN is currently employed by Pfizer; she contributed to this study before she became an employee of Pfizer. CSF is currently employed by Genentech, a subsidiary of Roche, and previously served as a consultant for Agios, Bain Capital, Bayer, Celgene, Dicerna, Eli Lilly, Entrinsic Health, Five Prime Therapeutics,

Genentech, Gilead Sciences, KEW, Merck, Merrimack Pharmaceuticals, Pfizer, Sanofi, Taiho and Unum Therapeutics. CSF also serves as a director for CytomX Therapeutics and owns unexercised stock options for CytomX and Entrinsic Health. ATC previously served as a consultant for Bayer Healthcare and Pfizer. MG receives research funding from Janssen and Sunbird Bio, consulting fees from Nerviano Medical Sciences, and honoraria from OncLive and PER. This study was not funded by any of these companies. No other conflicts of interest exist. The other authors declare that they have no conflicts of interest.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants. The study protocol was approved by the institutional review boards of Brigham and Women's Hospital and Harvard T.H. Chan School of Public Health (Boston, Massachusetts, USA; #2019P003588) and those of participating registries as required. Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request. With input from and approval of the External Advisory Committees for the two cohorts (the Nurses' Health Study (NHS) and the Health Professionals Follow-up Study (HPFS)), we have adopted a data enclave approach to data sharing. Along with a general description of the NHS and HPFS cohorts as well as all questionnaires that have been used since 1976 in the NHS and since 1986 in the HPFS, for both the blood repository and the questionnaire data, guidelines are available on our website for outside users to access the resources of our study (http://www.nurseshealthstudy.org/, http://www.hsph.harvard.edu/hpfs/). Typically, an outside user prepares a brief proposal that is reviewed by the NHS and HPFS investigator groups to identify a local investigator to assist the outside investigator. If uncertainty arises as to the merit of a request, the External Advisory Committees make a final decision.

ORCID iDs
Tsuyoshi Hamada http://orcid.org/0000-0002-3937-2755
Tomotaka Ugai http://orcid.org/0000-0003-0182-5269
Xuehong Zhang http://orcid.org/0000-0002-8260-8508
Koichiro Haruki http://orcid.org/0000-0002-1686-3228
Edward L Giovannucci http://orcid.org/0000-0002-6123-0219
Jonathan A Nowak http://orcid.org/0000-0002-0943-7407
Andrew T Chan http://orcid.org/0000-0001-7284-6767
Marios Giannakis http://orcid.org/0000-0001-9012-6982
Shuji Ogino http://orcid.org/0000-0002-3909-2323

REFERENCES

1 Lopez de Rodas M, Villalba-Esparza M, Sanmamed MF, *et al*. Biological and clinical significance of tumour-infiltrating lymphocytes in the era of immunotherapy: a multidimensional approach. *Nat Rev Clin Oncol* 2025;22:163–81.
2 Williams CJM, Peddle AM, Kasi PM, *et al*. Neoadjuvant immunotherapy for dMMR and pMMR colorectal cancers: therapeutic strategies and putative biomarkers of response. *Nat Rev Clin Oncol* 2024;21:839–51.
3 Park R, Saeed A. Immunotherapy in Colorectal Cancer - Finding the Achilles' Heel. *NEJM Evid* 2024;3.
4 André T, Elez E, Lenz H-J, *et al*. Nivolumab plus ipilimumab versus nivolumab in microsatellite instability-high metastatic colorectal cancer (CheckMate 8HW): a randomised, open-label, phase 3 trial. *Lancet* 2025;405:383–95.
5 Chalabi M, Verschoor YL, Tan PB, *et al*. Neoadjuvant Immunotherapy in Locally Advanced Mismatch Repair-Deficient Colon Cancer. *N Engl J Med* 2024;390:1949–58.
6 Giannakis M, Mu XJ, Shukla SA, *et al*. Genomic Correlates of Immune-Cell Infiltrates in Colorectal Carcinoma. *Cell Rep* 2016;15:857–65.
7 Yang K, Halima A, Chan TA. Antigen presentation in cancer — mechanisms and clinical implications for immunotherapy. *Nat Rev Clin Oncol* 2023;20:604–23.
8 Ghorani E, Swanton C, Quezada SA. Cancer cell-intrinsic mechanisms driving acquired immune tolerance. *Immunity* 2023;56:2270–95.
9 Westcott PMK, Muyas F, Hauck H, *et al*. Mismatch repair deficiency is not sufficient to elicit tumor immunogenicity. *Nat Genet* 2023;55:1686–95.
10 Budczies J, Kazdal D, Menzel M, *et al*. Tumour mutational burden: clinical utility, challenges and emerging improvements. *Nat Rev Clin Oncol* 2024;21:725–42.
11 Wang X, Lamberti G, Di Federico A, *et al*. Tumor mutational burden for the prediction of PD-(L)1 blockade efficacy in cancer: challenges and opportunities. *Ann Oncol* 2024;35:508–22.
12 Garofalo A, Sholl L, Reardon B, *et al*. The impact of tumor profiling approaches and genomic data strategies for cancer precision medicine. *Genome Med* 2016;8:79.
13 Amitay EL, Carr PR, Jansen L, *et al*. Smoking, alcohol consumption and colorectal cancer risk by molecular pathological subtypes and pathways. *Br J Cancer* 2020;122:1604–10.
14 Nishihara R, Morikawa T, Kuchiba A, *et al*. A prospective study of duration of smoking cessation and colorectal cancer risk by epigenetics-related tumor classification. *Am J Epidemiol* 2013;178:84–100.
15 Limsui D, Vierkant RA, Tillmans LS, *et al*. Cigarette smoking and colorectal cancer risk by molecularly defined subtypes. *J Natl Cancer Inst* 2010;102:1012–22.
16 Botteri E, Borroni E, Sloan EK, *et al*. Smoking and Colorectal Cancer Risk, Overall and by Molecular Subtypes: A Meta-Analysis. *Am J Gastroenterol* 2020;115:1940–9.
17 Alexandrov LB, Kim J, Haradhvala NJ, *et al*. The repertoire of mutational signatures in human cancer. *Nature New Biol* 2020;578:94–101.
18 Rizvi NA, Hellmann MD, Snyder A, *et al*. Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science* 2015;348:124–8.
19 Hamada T, Nowak JA, Masugi Y, *et al*. Smoking and Risk of Colorectal Cancer Sub-Classified by Tumor-Infiltrating T Cells. *J Natl Cancer Inst* 2019;111:42–51.
20 Ugai T, van Guelpen B, Mucci LA, *et al*. Enhancing existing tumour biobanks in European prospective cohort studies. *Lancet Reg Health Eur* 2025;53:101293.
21 Ogino S, Ugai T. The global epidemic of early-onset cancer: nature, nurture, or both? *Ann Oncol* 2024;35:1071–3.
22 Nishihara R, Wu K, Lochhead P, *et al*. Long-term colorectal-cancer incidence and mortality after lower endoscopy. *N Engl J Med* 2013;369:1095–105.
23 Borowsky J, Haruki K, Lau MC, *et al*. Association of *Fusobacterium nucleatum* with Specific T-cell Subsets in the Colorectal Carcinoma Microenvironment. *Clin Cancer Res* 2021;27:2816–26.
24 Ugai T, Väyrynen JP, Haruki K, *et al*. Smoking and Incidence of Colorectal Cancer Subclassified by Tumor-Associated Macrophage Infiltrates. *J Natl Cancer Inst* 2022;114:68–77.
25 Yamauchi M, Morikawa T, Kuchiba A, *et al*. Assessment of colorectal cancer molecular features along bowel subsites challenges the conception of distinct dichotomy of proximal versus distal colorectum. *Gut* 2012;61:847–54.
26 Haruki K, Kosumi K, Li P, *et al*. An integrated analysis of lymphocytic reaction, tumour molecular characteristics and patient survival in colorectal cancer. *Br J Cancer* 2020;122:1367–77.
27 Ogino S, Nosho K, Irahara N, *et al*. Lymphocytic reaction to colorectal cancer is associated with longer survival, independent of lymph node count, microsatellite instability, and CpG island methylator phenotype. *Clin Cancer Res* 2009;15:6412–20.
28 Ogino S, Kawasaki T, Kirkner GJ, *et al*. Evaluation of Markers for CpG Island Methylator Phenotype (CIMP) in Colorectal Cancer by a Large Population-Based Sample. *J Mol Diagn* 2007;9:305–14.
29 Ogino S, Kawasaki T, Brahmandam M, *et al*. Precision and performance characteristics of bisulfite conversion and real-time PCR (MethyLight) for quantitative DNA methylation analysis. *J Mol Diagn* 2006;8:209–17.

30 Morikawa T, Shima K, Kuchiba A, *et al*. No evidence for interference of h&e staining in DNA testing: usefulness of DNA extraction from H&E-stained archival tissue sections. *Am J Clin Pathol* 2012;138:122–9.
31 Gurjao C, Zhong R, Haruki K, *et al*. Discovery and Features of an Alkylating Signature in Colorectal Cancer. *Cancer Discov* 2021;11:2446–55.
32 Costello M, Pugh TJ, Fennell TJ, *et al*. Discovery and characterization of artifactual mutations in deep coverage targeted capture sequencing data due to oxidative DNA damage during sample preparation. *Nucleic Acids Res* 2013;41:e67.
33 Reynisson B, Alvarez B, Paul S, *et al*. NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Res* 2020;48:W449–54.
34 Shukla SA, Rooney MS, Rajasagi M, *et al*. Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes. *Nat Biotechnol* 2015;33:1152–8.
35 Aggarwal C, Ben-Shachar R, Gao Y, *et al*. Assessment of Tumor Mutational Burden and Outcomes in Patients With Diverse Advanced Cancers Treated With Immunotherapy. *JAMA Netw Open* 2023;6:e2311181.
36 Marcus L, Fashoyin-Aje LA, Donoghue M, *et al*. FDA Approval Summary: Pembrolizumab for the Treatment of Tumor Mutational Burden-High Solid Tumors. *Clin Cancer Res* 2021;27:4685–9.
37 Benjamin DJ, Berger JO, Johannesson M, *et al*. Redefine statistical significance. *Nat Hum Behav* 2018;2:6–10.
38 Lunn M, McNeil D. Applying Cox regression to competing risks. *Biometrics* 1995;51:524–32.
39 Wang M, Spiegelman D, Kuchiba A, *et al*. Statistical methods for studying disease subtype heterogeneity. *Stat Med* 2016;35:782–800.
40 Liu L, Nevo D, Nishihara R, *et al*. Utility of inverse probability weighting in molecular pathological epidemiology. *Eur J Epidemiol* 2018;33:381–92.
41 Inamura K, Hamada T, Bullman S, *et al*. Cancer as microenvironmental, systemic and environmental diseases: opportunity for transdisciplinary microbiomics science. *Gut* 2022;71:2107–22.
42 Ugai T, Sasamoto N, Lee H-Y, *et al*. Is early-onset cancer an emerging global epidemic? Current evidence and future implications. *Nat Rev Clin Oncol* 2022;19:656–73.
43 Akimoto N, Ugai T, Zhong R, *et al*. Rising incidence of early-onset colorectal cancer - a call to action. *Nat Rev Clin Oncol* 2021;18:230–43.
44 Lochhead P, Chan AT, Nishihara R, *et al*. Etiologic field effect: reappraisal of the field effect concept in cancer predisposition and progression. *Mod Pathol* 2015;28:14–29.
45 Jin Z, Zhou Q, Cheng J-N, *et al*. Heterogeneity of the tumor immune microenvironment and clinical interventions. *Front Med* 2023;17:617–48.
46 Bai J, Chen H, Bai X. Relationship between microsatellite status and immune microenvironment of colorectal cancer and its application to diagnosis and treatment. *J Clin Lab Anal* 2021;35:e23810.
47 Valero C, Lee M, Hoen D, *et al*. Response Rates to Anti-PD-1 Immunotherapy in Microsatellite-Stable Solid Tumors With 10 or More Mutations per Megabase. *JAMA Oncol* 2021;7:739–43.
48 Miao D, Margolis CA, Vokes NI, *et al*. Genomic correlates of response to immune checkpoint blockade in microsatellite-stable solid tumors. *Nat Genet* 2018;50:1271–81.
49 Hu L-F, Lan H-R, Huang D, *et al*. Personalized Immunotherapy in Colorectal Cancers: Where Do We Stand? *Front Oncol* 2021;11:769305.
50 Saint-André V, Charbit B, Biton A, *et al*. Smoking changes adaptive immunity with persistent effects. *Nature New Biol* 2024;626:827–35.
51 Grando SA. Connections of nicotine to cancer. *Nat Rev Cancer* 2014;14:419–29.
52 de la Iglesia JV, Slebos RJC, Martin-Gomez L, *et al*. Effects of Tobacco Smoking on the Tumor Immune Microenvironment in Head and Neck Squamous Cell Carcinoma. *Clin Cancer Res* 2020;26:1474–85.
53 Desrichard A, Kuo F, Chowell D, *et al*. Tobacco Smoking-Associated Alterations in the Immune Microenvironment of Squamous Cell Carcinomas. *J Natl Cancer Inst* 2018;110:1386–92.
54 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.
55 Ogino S, Chan AT, Fuchs CS, *et al*. Molecular pathological epidemiology of colorectal neoplasia: an emerging transdisciplinary and interdisciplinary field. *Gut* 2011;60:397–411.
56 Nakano S, Yamaji T, Shiraishi K, *et al*. Smoking and risk of colorectal cancer according to KRAS and BRAF mutation status in a Japanese prospective Study. *Carcinogenesis* 2023;44:476–84.
57 Cai J-A, Zhang Y-Z, Yu E-D, *et al*. Association of cigarette smoking with risk of colorectal cancer subtypes classified by gut microbiota. *Tob Induc Dis* 2023;21:99.
58 Hathaway CA, Wang T, Townsend MK, *et al*. Lifetime Exposure to Cigarette Smoke and Risk of Ovarian Cancer by T-cell Tumor Immune Infiltration. *Cancer Epidemiol Biomarkers Prev* 2023;32:66–73.
59 Sicklick JK, Kato S, Okamura R, *et al*. Molecular profiling of cancer patients enables personalized combination therapy: the I-PREDICT study. *Nat Med* 2019;25:744–50.
60 Sicklick JK, Kato S, Okamura R, *et al*. Molecular profiling of advanced malignancies guides first-line N-of-1 treatments in the I-PREDICT treatment-naïve study. *Genome Med* 2021;13:155.