Association between environmental modifiable risk score and molecular subtypes in colorectal cancer patients

Mary Grace Casagrande

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Public Health

University of Washington
2023

Committee:
Ulrike Peters
Amanda Phipps

Program Authorized to Offer Degree:
Department of Epidemiology

University of Washington

Abstract

Association between environmental modifiable risk score and molecular subtypes in colorectal cancer patients

Mary Grace Casagrande

Chair of the Supervisory Committee:
Ulrike Peters
Department of Epidemiology

Background: Colorectal cancer (CRC) is associated with several modifiable environmental risks, including dietary and non-dietary factors like calcium intake and smoking habits. It is unknown if the association between overall environmental risk and CRC is differentially associated with certain molecular subtypes.

Methods: A total of 6389 cases and 6835 controls of European ancestry from 11 observational studies were included. We combined 13 modifiable risk factors of CRC to create an Environmental Risk Score (ERS) that characterizes overall environmental predisposition to CRC. Four somatic colorectal tumor markers were assessed individually and in combination, including CpG Island Methylator phenotype (CIMP), oncogenic mutations in *KRAS* and *BRAF*, and the presence of microsatellite instability (MSI). A multinomial logistic regression was used to assess the association between ERS and risk of CRC molecular subtypes, adjusting for age, sex, study, and energy intake. We also stratified analyses by sex as a sensitivity analysis.

Results: CRC risk was positively associated with the ERS [odds ratio (OR)=2.67 (95% CI: 2.40, 2.97)]. Associations between ERS and CRC risk were modestly stronger for *KRAS*-wildtype CRC compared to *KRAS*-mutated CRC ($P_{difference}$ = 0.063), particularly in females ($P_{difference}$ = 0.029). Associations between ERS and CRC risk were consistent in regard to *BRAF* mutation,

CIMP, and MSI status. All molecular subtypes defined by combinations of tumor markers showed uniformity in their association with the ERS and CRC risk, although Jass type 3 tumors (MSI-low/stable, CIMP-low/negative, *BRAF*-wildtype, *KRAS*-mutated) in females and Jass type 5 tumors (MSI-high, CIMP-low/negative, *BRAF*-wildtype, *KRAS*-wildtype) in males differed significantly in their association with the ERS from the predominant Jass type 4 ($P_{difference}$ = 0.044 and 0.015, respectively).

Discussion: The ERS was strongly associated with CRC risk overall and across most subgroups of CRC defined by tumor characteristics. Our results suggest that *KRAS*-wildtype tumors are more strongly associated with environmental factors than are *KRAS*-mutated tumors, particularly in women.

I. Introduction

Colorectal cancer (CRC) is the third most common cancer in both males and females in countries with a high Human Development Index (HDI) (1). Four of the most predominant tumor markers that have been studied in CRC are CpG Island Methylator phenotype (CIMP), oncogenic mutations in *KRAS* and *BRAF*, and the presence of microsatellite instability (MSI) (likely reflecting deficiencies in DNA mismatch repair) (2). It is estimated that 60-65% of CRC cases arise without a known family history of CRC or inherited genetic mutation known to be related to CRC (3). It is estimated that between 15% to 35% of the variation in CRC risk is contributed by genetic factors, with possibly up to 5% estimated to be caused by hereditary cancer syndromes like Lynch syndrome (1–3). However, even for the cases of CRC caused by heritable factors, most are not solely due to genetics; there are also several environmental factors that contribute to the incidence of CRC (3).

There are many modifiable environmental risk factors of CRC, several of which are related to diet. There is a multitude of evidence that high intake of red meat, and processed meat in particular, can contribute to CRC (4). One meta-analysis showed that eating 100g of red meat per day or 50g of processed meat per day can lead to a 15-20% increased risk of CRC (4). Alcohol is associated with an increased risk of CRC, particularly in men (4). One study showed that consuming 30g of alcohol per day can increase CRC risk by 16% (4). Another meta-analysis using the same dataset used in this study found that consuming less than 1 drink (defined as 14 g) of alcohol per day had a protective effect against CRC and consuming 3 or more drinks was associated with an increased risk of CRC (4,5). There are also dietary factors that show a protective effect against CRC. Dietary fiber has an association with protection against CRC (4). One study showed that risk of CRC can decrease up to 25% when consuming between 12.6g and 33.1g of dietary fiber per day (4). Fruits, vegetables, and calcium have been shown to have small protective effects (4). Adoption of a “Western” diet (high intake of red meat and/or processed meat, high-fat dairy products, fast food, refined grains, and sweet foods and drinks) is associated with an increased risk of CRC (6).

Other environmental factors linked to CRC risk include smoking, physical inactivity, and obesity (7). Smoking has been shown to lead to increased risk of CRC, particularly in women, with one meta-analysis showing that smoking can increase risk of CRC by 15-20% (8). Obesity and low levels of physical activity are highly associated with risk of CRC, with individuals with a body mass index (BMI) > 30 kg/m^2 showing a 19% increased risk of CRC as compared to those with a BMI between 20 and 25 kg/m^2, and individuals who regularly exercise decreasing their CRC risk by 40% compared to those who do not (4,9). There are also protective medications. Those who take post-menopausal hormone therapy (PMH) have been shown to have a lower risk of CRC (10). Non-steroidal anti-inflammatory (NSAID) drugs, such as aspirin, have been shown to reduce risk of CRC (7).

Some of these factors have also been associated with decreased or increased risk of CRC exhibiting specific tumor markers. Smoking has been highly associated with CIMP-High status, *BRAF* mutations, and MSI-High status in CRC (7,8). NSAIDs have been associated with

protection against CRC tumors with *KRAS* or *BRAF* mutations (7). Calcium intake has been shown to have a small protective effect against CIMP-high and MSI-high tumors (7). Obesity is strongly associated with microsatellite stable (MSS) or MSI-low tumors (7).

These tumor markers have been shown to group together in particular molecular subtypes, also known as Jass types, reflecting distinct etiologies (11). Type 2 tumors (CIMP-high / MSI-low or MSS / *BRAF*-mutated) have been shown to have the highest mortality and type 5 (CIMP-negative / MSI-high / *BRAF* and *KRAS* wildtype) has been shown to have the lowest mortality, while type 4 tumors (CIMP-negative / MSS or MSI-low / *BRAF* and *KRAS* wildtype) are the most common (12,13).

The environmental risk score (ERS) used in this study is based on an aggregate of environmental risk factors to assess risk. Using an ERS, as opposed to the individual risk factors, can be useful because it allows us to evaluate an individual's overall lifestyle and aggregate risk factor profile rather than just one factor alone. Such scores have been used in the past for similar studies (14,15). In one instance, a more healthy lifestyle was associated with overall lower risk of CRC, even if patients had a high genetic risk (16). In another study, a score was generated from just four factors, and a high score was associated with a higher risk of CRC (17). Other ERS's similar to the one used in this study have been used to determine risk and optimal age at screening initiation (15).

This study seeks to identify whether or not there is an association between the presence of the tumor markers mentioned above (individually or in combination) and the level of modifiable environmental risk in both CRC cases and controls.

II. Methods

Data was obtained from the Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO) and the Colon Cancer Family Registry (CCFR), which are international collaborations that focus on the identification and characterization of genetic risk factors and gene-environment interactions for CRC (18). GECCO and CCFR further aim to analyze molecular and genetic characteristics of colorectal tumors and their microenvironment in relation to host genomes, lifestyle and environmental risk factors, and survival. This study included data collected from CCFR, Cancer Prevention Study II (CPSII), Darmkrebs: Chancen der Verhütung durch Screening (DACHS), Diet, Activity, and Lifestyle Study (DALS), Early Detection Research Network (EDRN), European Prospective Investigation into Cancer (Sweden) (EPIC), Health Professionals Follow-up Study (HPFS), Melbourne Collaborative Cohort Study (MCCS), Newfoundland Familial Colorectal Cancer Study (NFCCR), Nurses' Health Study (NHS), and Northern Swedish Health and Disease Study (NSHDS). The dataset includes a total of 23,751 individuals. The CRC case-only analysis includes 7,454 individuals, with 3,911 females and 3,543 males. The case-control analyses include 13,224 individuals, with 7,198 females (3,388 cases and 3,810 controls) and 6,026 males (3,001 cases and 3,025 controls). (See Table 1 for details on characteristics of individual studies.) Only individuals of European descent were included due to insufficient sample size for other racial and ethnic groups. All participants gave

written informed consent, and studies were approved by their respective Institutional Review Boards.

Table 1- *Characteristics of Individual Studies in GECCO*

Study	**Country**	**Study design**	**Years of incident CRC diagnoses**	**Number of cases in dataset**	**Number of controls in dataset**
CCFR	USA, Canada and Australia	Consortium	1998-2011	5045	2469
CPSII	USA	Cohort	1984-2006	860	1003
DACHS	Germany	Case-Control	2003-2014	2009	2789
DALS	USA	Case-Control	1991-1994	1095	1162
EDRN	USA	Consortium	2000-2013	219	356
EPIC-Sweden	Sweden	Cohort	2016	146	381
HPFS	USA	Cohort	1986-2011	626	602
MCCS	Australia	Cohort	1990-1994	490	674
NFCCR	Canada	Case-Control	1999 and 2003	573	472
NHS	USA	Cohort	1976-2013	793	1242
NSHDS	Sweden	Cohort	1986-2008	327	415

a. Environmental Risk Score

Demographic, behavioral, diet and pharmacological factors were collected via in-person interviews and structured questionnaires (15,16). We developed a combined environmental risk score (ERS) based on 12 modifiable factors selected based on known and plausible CRC risk factors previously reported in the literature: intake of red meat, intake of processed meat, alcohol intake, dietary fiber intake, calcium intake, fruit intake, vegetable intake, smoking status (defined as ever smoked vs never smoked and pack-years), aspirin/NSAID intake, PMH status, level of physical activity, and BMI. To account for differences in the assessment tools between studies and sex, we coded dietary variables and pack-years smoked into sex- and study-specific quartiles, with quartile cutoffs determined within the controls of each study and sex. Alcohol was

coded as: non-/occasional drinkers (drinking < 1 g/day); light-to-moderate drinkers (drinking 1–28 g/day); and heavy drinkers (drinking > 28 g/d). Physical activity was defined as >1 hour/week (active) and < 1 hour/week (inactive). The aspirin/NSAID and PMH variables were coded as yes/no for regular use. BMI was coded per 5 units of increase (kg/m2). 204 patients were dropped from the EDRN study because their alcohol data was missing.

To develop the ERS, we used a multivariable logistic regression model that includes all risk factors and adjusts for age, energy intake, sex, and study. Then each subject's value was multiplied by the regression coefficient for each variable and summed across variables. The ERS score was then coded into quartiles.

b. Tumor Subtypes and Outcomes

The presence of four tumor markers (*BRAF* oncogenic mutations, *KRAS* oncogenic mutations, CIMP status, and MSI status) was determined by obtaining formalin-fixed paraffin-embedded (FFPE) tumor tissue as described elsewhere (13). Individual markers were classified based on the presence vs. absence of each tumor attribute (i.e., *BRAF*-wildtype/*BRAF*-mutated, *KRAS*-wildtype/*KRAS*-mutated, MSI-high / MSS or MSI-low, CIMP-high / CIMP-low or -negative). Tumor subtypes were defined based on tumor marker combinations consistent with previously described Jass tumor subtypes: Jass type 1 is CIMP-high/MSI-high/ *BRAF*-mutated; Jass type 2 is CIMP-high/MSI-low or MSS/ *BRAF*-mutated; Jass type 3 is CIMP-low/ MSS or MSI-low/*KRAS-mutated*. Jass type 4 is CIMP-negative/MSS or MSI-low/*BRAF* and *KRAS* wildtype; Jass type 5 is CIMP-negative/ MSI-high/*BRAF* and *KRAS* wildtype (11,13).

c. Statistical Analyses

Model covariates for case-only and case-control analyses included a study population identifier and age, as defined by the age at which the exposure variables were assessed. Energy level was also assessed as a covariate and defined as kcal/day. Due to the presence of variables that are sex-specific, such as PMH, and the differing levels of risk by sex as determined by previous studies, analyses were stratified by sex, defined as male and female, as a sensitivity analysis (15,16,19). For missing values of ERS variables, the sex study-specific means were imputed and substituted. For the overall analysis, the males had their PMH values substituted with the mean value for females.

Analyses were conducted in four stages: 1) overall CRC risk and environmental risk (e.g., case-control status and ERS score), 2) case-only analyses (e.g., MSI-high vs. MSS/MSI-low cases) to investigate if associations of ERS and CRC differ significantly by CRC subtype (assessment of p for heterogeneity), 3) individual marker case-control analysis (e.g., MSI-high cases vs. controls, MSS/MSI-low cases vs. controls), and 4) aggregate subtype case-control analysis (e.g. Jass type 1 vs. controls). For case-only analyses, p-values were generated based on a Wald test from a multivariable logistic regression model. For the case-control analyses, we estimated odds ratios (ORs) and 95% confidence intervals (CIs) from polytomous multivariable logistic regression models to quantify the association between the ERS and each individual tumor marker as well as each Jass type. All models were also stratified by sex to allow for

variation in risk factor effects for men and women, as a sensitivity analysis. All significance was assessed at $\alpha = 0.05$. All analyses were conducted using R V.4.2.2 (http://www-r-project.org).

III. Results

a. Descriptive Results

For both the sex-stratified and the sex-combined analysis, cases were less likely to have participated in physical activity and less likely to have regularly used NSAIDs or aspirin than controls (Table 2). Also, female cases were less likely to have used PMH than controls. Cases had a lower mean intake of processed meat and calcium than controls, and controls had a higher mean energy intake. All of these variables were included in the ERS, along with several others.

Table 2- Descriptive Characteristics of Pharmacological, Demographic, Dietary and Lifestyle Risk Factors in Study Population

Variable	Males		Females		Overall	
	Cases (n = 3001)	Controls (n = 3025)	Cases (n =3388)	Controls (n=3810)	Cases (n= 6839)	Controls (n= 6835)
Age, years, mean (SD)	62.30 (11.66)	64.82 (11.33)	62.83 (11.82)	64.20 (10.81)	62.58 (11.75)	64.47 (11.05)
BMI, (kg/m2)/5, mean (SD)	5.52 (0.84)	5.36 (0.71)	5.34 (1.08)	5.16 (0.94)	5.43 (0.98)	5.25 (0.86)
Physical activity, yes, n (%)	1575 (50.38)	2214 (73.19)	1238 (36.54)	1611 (42.28)	3825 (44.0)	2813 (56.0)
Smoking status, ever, n (%)	1837 (61.21)	1829 (60.46)	1545 (45.60)	1555 (40.81)	3384 (52.9)	3382 (49.5)
Smoking, pack-years, mean (SD)	1.68 (1.50)	1.48 (1.45)	1.20 (1.51)	1.01 (1.40)	1.43 (1.52)	1.22 (1.44)
Alcohol consumptio						

n, yes, n (%)						
<1 g/d	781 (26.02)	721 (23.83)	1697 (50.09)	1641 (43.07)	2478 (38.8)	2362 (34.6)
1–28 g/d	1426 (47.52)	1615 (53.39)	1453 (42.89)	1901 (49.90)	2879 (45.1)	3516 (51.4)
>28 g/d	794 (26.46)	689 (22.78)	238 (7.02)	268 (7.03)	1032 (16.2)	957 (14.0)
Aspirin/ NSAID use, yes, n (%)	794 (26.46)	997 (32.96)	875 (23.85)	1319 (34.46)	1669 (26.1)	2316 (33.9)
PMH use, yes, n (%)			959 (28.31)	1485 (38.98)		
Fiber Intake, mean quartile (SD)	0.66 (1.23)	0.84 (1.35)	1.19 (1.43)	1.51 (1.46)	0.94 (1.37)	1.21 (1.45)
Red Meat Intake, mean quartile (SD)	2.42 (1.11)	2.23 (0.96)	2.41 (1.04)	2.40 (1.04)	2.42 (1.03)	2.33 (1.01)
Processed Meat Intake, mean quartile (SD)	1.57 (1.40)	1.96 (1.16)	1.76 (1.41)	2.10 (1.25)	1.67 (1.40)	2.04 (1.21)
Fruit Intake, mean quartile (SD)	2.26 (0.71)	2.28 (0.70)	2.31 (0.89)	2.41 (0.86)	2.29 (0.81)	2.35 (0.79)

Vegetable Intake, mean quartile (SD)	2.29 (0.89)	2.30 (0.79)	2.38 (0.78)	2.42 (0.83)	2.34 (0.83)	2.37 (0.81)
Calcium Intake, mean quartile (SD)	2.39 (0.96)	2.48 (1.07)	2.36 (0.94)	2.46 (1.03)	2.37 (0.95)	2.47 (1.05)
Energy Intake, kcal/day, mean (SD)[1]	2093 (700.3)	2090 (690.6)	1667 (533.0)	1685 (520.9)	1808 (626.7)	1806 (606.1)
ERS, mean (SD)	1.25 (0.38)	1.15 (0.37)	0.37 (0.40)	0.23 (0.40)	0.63 (0.37)	0.50 (0.38)
ERS quartile, mean (SD)	2.71 (1.09)	2.29 (1.10)	2.68 (1.10)	2.34 (1.11)	2.69 (1.09)	2.33 (1.12)

1 = CCFR-Seattle, CCFR-Australia, and DACHS studies did not measure total energy input and were not included in this calculation

KRAS-mutated tumors were the most common of the individual molecular subtypes and Jass type 4 tumors were the most common of the aggregate molecular subtypes. Females were slightly more likely to have tumors that exhibited individual tumor markers than males, and were more likely to have tumors exhibiting Jass type 1, 2, 3, and 5 phenotypes. Males were more likely to have Jass type 4 phenotypes (see supplementary table 1).

Supplementary Table 1

Tumor subtype/ Jass type	Male Cases	Female Cases	Overall Cases
BRAF-mutated, n (%*)	210 (7.59)	536 (17.05)	746 (12.6)

KRAS-mutated, n (%*)	790 (30.49)	949 (31.88)	1739 (31.2)
CIMP-high, n (%*)	320 (12.05)	687 (22.73)	1007 (17.7)
MSI-high, n (%*)	265 (9.75)	553 (17.72)	818 (14.00)
Jass type 1+, n (%*)	57 (3.08)	265 (12.11)	322 (8.00)
Jass type 2+, n (%*)	39 (2.11)	104 (4.75)	143 (3.54)
Jass type 3+, n (%*)	522 (28.20)	704 (18.60)	1226, (30.35)
Jass type 4+, n (%*)	1174 (63.42)	1046 (47.81)	2220 (55.00)
Jass type 5+, n (%*)	59 (3.19)	69 (3.15)	128 (3.17)

**= percentage of cases, not full population*

b. Analytic Results

We observed a strong positive association between ERS score and CRC risk overall OR comparing the 4th vs the 1st quartile: 2.67 (95% CI: 2.40, 2.97) which was very similar among females and males (female OR: 2.72 (95% CI: 2.38, 3.11), male OR: 2.72 (95% CI: 2.33, 3.18)). Given these similarities, we conducted sex-combined analysis, and conducted the sex-stratified analysis as a sensitivity analysis (see supplementary table 2 for overall CRC risk results).

Supplementary Table 2: Overall risk assessed for ERS

	ERS OR	95% CI
Overall (including PMH)	2.67	2.40, 2.97
Female	2.72	2.38, 3.11
Male	2.72	2.33, 3.18

Associations between increasing ERS and increasing CRC risk were largely consistent regardless of *BRAF*, MSI, or CIMP status (p = 0.70, 0.67, 0.71, respectively). Associations were modestly stronger with respect to risk of *KRAS*-wildtype CRC compared with *KRAS*-mutated CRC; however, this difference was not statistically significant (p-heterogeneity=0.063) (Table 3).

Table 3: Odds ratios (95% confidence intervals) for the association of Environmental Risk Score with CRC Risk by *KRAS, BRAF,* microsatellite and CIMP status

KRAS	Quartile	*KRAS*-mutated Cases (N, %)	*KRAS*-wildtype Cases (N, %)	Controls (N, %)	*KRAS*-mutated vs Controls, adj OR (95% CI)	*KRAS*-wildtype vs Controls, adj OR (95% CI)	P Heterogeneity (case/case analysis)
	Lowest (Q1)	351 (20.2)	692 (18.1)	2126 (31.1)	1 (ref)	1 (ref)	
	Second (Q2)	444 (25.5)	901 (23.5)	1766 (25.8)	1.54 (1.32, 1.80)	1.58 (1.40, 1.78)	
	Third (Q3)	458 (26.3)	1071 (28.0)	1537 (22.5)	1.89 (1.61, 2.22)	2.20 (1.94, 2.48)	
	Highest (Q4)	486 (27.9)	1165 (30.4)	1406 (20.6)	2.30 (1.94, 2.72)	2.71 (2.38, 3.09)	
	P_{trend}				6.58e-06	1.25e-13	0.063
BRAF	Quartile	*BRAF*-mutated Cases (N, %)	*BRAF*-wildtype Cases (N, %)	Controls (N, %)	*BRAF*-mutated vs Controls, adj OR (95% CI)	*BRAF*-wildtype vs Controls, adj OR (95% CI)	P Heterogeneity (case/case analysis)

	Lowest (Q1)	140 (18.8)	935 (18.1)	2126 (31.1)	1 (ref)	1 (ref)	
	Second (Q2)	164 (22.0)	1239 (24.1)	1766 (25.8)	1.47 (1.16, 1.87)	1.59 (1.42, 1.78)	
	Third(Q3)	199 (26.7)	1448 (28.0)	1537 (22.5)	1.98 (1.56, 2.50)	2.17 (1.93, 2.42)	
	Highest (Q4)	243 (32.6)	1541 (29.8)	1406 (20.6)	2.40 (1.89, 3.05)	2.70 (2.40, 3.05)	
	P_{trend}				1.01e-03	1.33e-15	0.70
MSI	Quartile	MSI-high Cases (N, %)	MSI-low or MSS Cases (N, %)	Controls (N, %)	MSI-high vs Controls, adj OR (95% CI)	MSI-low or MSS vs Controls, adj OR (95% CI)	P Heterogeneity (case/case analysis)
	Lowest (Q1)	158 (19.3)	903 (18.0)	2126 (31.1)	1 (ref)	1 (ref)	
	Second (Q2)	193 (23.6)	1207 (24.0)	1766 (25.8)	1.53 (1.22, 1.91)	1.60 (1.44, 1.79)	
	Third(Q3)	211 (25.8)	1398 (27.8)	1537 (22.5)	1.94 (1.55, 2.43)	2.14 (1.91, 2.40)	
	Highest (Q4)	256 (31.3)	1513 (30.1)	1406 (20.6)	2.49 (1.98, 3.13)	2.66 (2.36, 3.01)	
	P_{trend}				3.44e-04	9.10e-15	0.65

CIMP	Quartile	CIMP -high Cases (N, %)	CIMP- low or negative Cases (N, %)	Controls (N, %)	CIMP- high vs Controls, adj OR (95% CI)	CIMP- low or negative vs Controls, adj OR (95% CI)	P Heterogeneity (case/case analysis)
	Lowest (Q1)	203 (20.2)	876 (18.8)	2126 (31.1)	1 (ref)	1 (ref)	
	Second (Q2)	263 (26.1)	1121 (24.0)	1766 (25.8)	1.75 (1.43, 2.13)	1.55 (1.39, 1.73)	
	Third(Q3)	268 (26.6)	1302 (27.9)	1537 (22.5)	2.23 (1.82, 2.73)	2.13 (1.89, 2.38)	
	Highest (Q4)	273 (27.1)	1371 (29.4)	1406 (20.6)	2.57 (2.08, 3.18)	2.62 (2.32, 2.96)	
	P_{trend}				5.03e-05	2.53e-14	0.71

Patterns of association between ERS and CRC risk according to *KRAS* mutation, *BRAF* mutation, CIMP, and MSI status were similar among males and among females, with modestly stronger associations observed for *KRAS*-wildtype tumors (Females Q4: *KRAS*-mutated- 2.12 (1.70, 2.66), *KRAS*-wildtype-2.87 (2.42, 3.42) p-heterogeneity=0.029; Males Q4: *KRAS* mutated- 2.29 (1.80, 2.92), *KRAS*-wildtype- 2.63 (2.19, 3.16), p-heterogeneity=0.06) (supplementary table 3).

Supplementary Table 3: Odds ratios (95% confidence intervals) for the association of Environmental Risk Score with CRC Risk by *KRAS, BRAF,* microsatellite and CIMP status, stratified by gender

KRAS	Males	Quartile	*KRAS*-mutated Cases (N, %)	*KRAS*-wildtype Cases (N, %)	Controls (N, %)	*KRAS*-mutated vs Controls, adj OR (95% CI)	*KRAS*-wildtype vs Controls, adj OR (95% CI)	P Heterogeneity (case/case analysis)
		Lowest (Q1)	155 (19.6)	327 (18.2)	960 (31.6)	1 (ref)	1 (ref)	
		Second (Q2)	196 (24.8)	414 (23.0)	797 (26.3)	1.44 (1.14, 1.82)	1.42 (1.18, 1.69)	
		Third (Q3)	213 (27.0)	478 (26.5)	702 (23.2)	1.76 (1.39, 2.22)	1.80 (1.50, 2.15)	
		Highest (Q4)	226 (28.6)	582 (32.2)	570 (18.8)	2.29 (1.80, 2.92)	2.63 (2.19, 3.16)	
		P_{trend}				3.22e-03	1.46e-06	0.06
	Females	Quartile	*KRAS*-mutated Cases (N, %)	*KRAS*-wildtype (N, %)	Controls (N, %)	*KRAS*-mutated vs Controls, adj OR (95% CI)	*KRAS*-wildtype - vs Controls, adj OR (95% CI)	P Heterogeneity
		Lowest (Q1)	201 (21.2)	382 (18.8)	1139 (29.9)	1 (ref)	1 (ref)	
		Second (Q2)	231 (24.3)	467 (23.0)	1002 (26.3)	1.37 (1.11, 1.69)	1.49 (1.27, 1.76)	

		Third (Q3)	258 (27.2)	533 (26.3)	901 (23.6)	1.80 (1.45, 2.24)	1.99 (1.68, 2.35)	
		Highest (Q4)	259 (27.3)	646 (31.9)	768 (20.2)	2.12 (1.70, 2.66)	2.87 (2.42, 3.42)	
		P_{trend}				2.87e-03	1.74e-08	0.029
BRAF	Males	Quartile	*BRAF*-mutated Cases (N, %)	*BRAF*-wildtype Cases (N, %)	Controls (N, %)	*BRAF*-mutated vs Controls, adj OR (95% CI)	*BRAF*-wildtype vs Controls, adj OR (95% CI)	P Heterogeneity (case/case analysis)
		Lowest (Q1)	36 (17.1)	462 (18.1)	960 (31.6)	1 (ref)	1 (ref)	
		Second (Q2)	54 (25.7)	591 (23.1)	797 (26.3)	1.69 (1.09, 2.62)	1.43 (1.22, 1.67)	
		Third (Q3)	61 (29.0)	682 (26.7)	702 (23.2)	1.93 (1.24, 2.99)	1.82 (1.55, 2.14)	
		Highest (Q4)	59 (28.1)	820 (32.1)	570 (18.8)	2.14 (1.36, 3.36)	2.68 (2.27, 3.17)	
		P_{trend}				0.26	5.20e-08	0.60

Females	Quartile	*BRAF*-mutated Cases (N, %)	*BRAF*-wildtype Cases (N, %)	Controls (N, %)	*BRAF*-mutated vs Controls, adj OR (95% CI)	*BRAF*-wildtype vs Controls, adj OR (95% CI)	P Heterogeneity (case/case analysis)
	Lowest (Q1)	100 (18.7)	496 (19.0)	1139 (29.9)	1 (ref)	1 (ref)	
	Second (Q2)	121 (22.6)	619 (23.7)	1002 (26.3)	1.38 (1.04, 1.84)	1.52 (1.31, 1.77)	
	Third (Q3)	131 (24.4)	704 (27.0)	901 (23.6)	1.62 (1.21, 2.16)	2.04 (1.74, 2.39)	
	Highest (Q4)	184 (34.3)	789 (30.3)	768 (20.2)	2.74 (1.95, 3.42)	2.74 (2.33, 3.22)	
	P_{trend}				4.03e-03	9.40e-09	0.94

MSI	Males	Quartile	MSI-high Cases (N, %)	MSI-low or MSS Cases (N, %)	Controls (N, %)	MSI-high vs Controls, adj OR (95% CI)	MSI-low or MSS vs Controls, adj OR (95% CI)	P Heterogeneity (case/case analysis)
		Lowest (Q1)	51 (19.2)	435 (17.7)	960 (31.6)	1 (ref)	1 (ref)	
		Second (Q2)	73 (27.5)	556 (22.7)	797 (26.3)	1.60 (1.10, 2.32)	1.42 (1.21, 1.67)	
		Third (Q3)	66 (24.9)	674 (27.5)	702 (23.2)	1.55 (1.05, 2.29)	1.91 (1.62, 2.24)	

		Highest (Q4)	75 (28.3)	789 (32.2)	570 (18.8)	2.09 (1.42, 3.09)	2.68 (2.26, 3.17)	
		P_{trend}				0.22	4.96e-08	0.18
	Females	Quartile	MSI-high Cases (N, %)	MSI-low or MSS Cases (N, %)	Controls (N, %)	MSI-high vs Controls, adj OR (95% CI)	MSI-low or MSS vs Controls, adj OR (95% CI)	P Heterogeneity (case/case analysis)
		Lowest (Q1)	88 (15.9)	508 (19.8)	1139 (29.9)	1 (ref)	1 (ref)	
		Second (Q2)	144 (26.0)	585 (22.8)	1002 (26.3)	1.94 (1.46, 2.58)	1.39 (1.20, 1.62)	
		Third (Q3)	131 (23.7)	705 (27.5)	901 (23.6)	2.03 (1.51, 2.73)	1.95 (1.67, 2.28)	
		Highest (Q4)	190 (34.4)	769 (30.0)	768 (20.3)	3.37 (2.52, 4.50)	2.54 (2.16, 2.98)	
		P_{trend}				4.45e-04	6.53e-08	0.67
CIMP	Males	Quartile	CIMP-high Cases (N, %)	CIMP-low or negative Cases (N, %)	Controls (N, %)	CIMP-high vs Controls, adj OR (95% CI)	CIMP-low or negative vs Controls, adj OR (95% CI)	P Heterogeneity (case/case analysis)

		Lowest (Q1)	58 (18.1)	445 (19.1)	960 (31.6)	1 (ref)	1 (ref)	
		Second (Q2)	92 (28.8)	534 (22.9)	797 (26.3)	2.09 (1.48, 2.96)	1.36 (1.15, 1.60)	
		Third (Q3)	81 (25.3)	627 (26.9)	702 (23.2)	2.13 (1.49, 3.05)	1.79 (1.52, 2.10)	
		Highest (Q4)	89 (27.8)	729 (31.2)	570 (18.8)	3.20 (2.22, 4.60)	2.54 (2.14, 3.00)	
		P_{trend}				0.012	2.94e-07	0.84
	Females	Quartile	CIMP-high Cases (N, %)	CIMP-low or negative Cases (N, %)	Controls (N, %)	CIMP-high vs Controls, adj OR (95% CI)	CIMP-low or negative vs Controls, adj OR (95% CI)	P Heterogeneity (case/case analysis)
		Lowest (Q1)	141 (20.5)	460 (19.7)	1139 (29.9)	1 (ref)	1 (ref)	
		Second (Q2)	179 (26.1)	544 (23.3)	1002 (26.3)	1.56 (1.22, 1.98)	1.46 (1.25, 1.71)	
		Third (Q3)	164 (23.9)	638 (27.3)	901 (23.6)	1.70 (1.31, 2.19)	2.03 (1.73, 2.39)	
		Highest (Q4)	203 (29.5)	693 (29.7)	768 (20.2)	2.51 (1.95, 3.24)	2.67 (2.26, 3.16)	
		P_{trend}				2.92e-03	3.37e-08	0.29

Patterns of association between the ERS and CRC risk were similarly consistent across Jass molecular subtypes (Table 4).

Table 4: Odds ratios (95% confidence intervals) for the association of Environmental Risk Score with CRC Risk by Jass type status

	Quartile	Jass Type Cases (N, %)	Controls (N, %)	Jass type vs Controls, adj OR (95% CI)	P Heterogeneity (case/case analysis)
Jass type 1	Lowest (Q1)	68 (21.1)	2126 (31.1)	1 (ref)	
	Second (Q2)	71 (22.0)	1766 (25.8)	1.37 (0.97, 1.94)	
	Third(Q3)	79 (24.5)	1537 (22.5)	1.82 (1.28, 2.57)	
	Highest (Q4)	104 (32.3)	1406 (20.6)	2.32 (1.65, 3.28)	
	P_{trend}			0.11	0.58
Jass type 2	Lowest (Q1)	20 (14.0)	2126 (31.1)	1 (ref)	
	Second (Q2)	35 (24.5)	1766 (25.8)	2.05 (1.17, 3.59)	
	Third(Q3)	46 (32.2)	1537 (22.5)	2.86 (1.65, 4.96)	
	Highest (Q4)	42 (29.4)	1406 (20.6)	2.45 (1.38, 4.35)	
	P_{trend}			0.34	0.32
Jass type 3	Lowest (Q1)	248 (20.2)	2126 (31.1)	1 (ref)	
	Second (Q2)	298 (24.3)	1766 (25.8)	1.45 (1.21, 1.74)	

	Third(Q3)	316 (25.8)	1537 (22.5)	1.83 (1.51, 2.20)	
	Highest (Q4)	364 (29.7)	1406 (20.6)	2.32 (1.92, 2.82)	
	P_{trend}			0.00011	0.07
Jass type 4	Lowest (Q1)	395 (17.8)	2126 (31.1)	1 (ref)	
	Second (Q2)	505 (22.7)	1766 (25.8)	1.53 (1.32, 1.78)	
	Third(Q3)	631 (28.4)	1537 (22.5)	2.27 (1.96, 2.64)	
	Highest (Q4)	689 (31.0)	1406 (20.6)	2.92 (2.49, 3.41)	
	P_{trend}			1.29e-11	ref
Jass type 5	Lowest (Q1)	24 (18.8)	2126 (31.1)	1 (ref)	
	Second (Q2)	29 (22.7)	1766 (25.8)	1.33 (0.92, 2.42)	
	Third(Q3)	41 (32.0)	1537 (22.5)	2.06 (1.34, 3.90)	
	Highest (Q4)	34 (26.6)	1406 (20.6)	1.83 (1.04, 3.32)	
	P_{trend}			0.46	0.06

In analyses stratified by sex, the ERS was not associated with risk of Jass type 5 tumors in males or females; however, these analyses were limited by small numbers (supplementary table 4). For females, the ERS was slightly less strongly associated with risk of Jass type 3 (Q4: 2.17 (1.69, 2.80)) tumors than with risk of Jass type 4 CRC (Q4: 3.04 (2.44, 3.79), p = 0.044).

Supplementary Table 4: Odds ratios (95% confidence intervals) for the association of Environmental Risk Score with CRC Risk by Jass type status, differentiated by gender

		Quartile	Jass type Cases (N, %)	Controls (N, %)	Jass type vs Controls, adj OR (95% CI)	P Heterogeneity (case/case analysis)
Jass type 1	Females	Lowest (Q1)	53 (20.0)	1139 (29.9)	1 (ref)	
		Second (Q2)	61 (23.0)	1002 (26.3)	1.30 (0.88, 1.91)	
		Third (Q3)	56 (21.1)	901 (23.6)	1.29 (0.86, 1.94)	
		Highest (Q4)	95 (35.8)	768 (20.2)	2.48 (1.70, 3.62)	
		P_{trend}			0.0728	0.31
	Males	Lowest (Q1)	10 (17.5)	879 (29.1)	1 (ref)	
		Second (Q2)	18 (31.6)	775 (25.6)	2.29 (1.04, 5.05)	
		Third(Q3)	16 (28.1)	728 (24.1)	2.25 (0.99, 5.11)	
		Highest (Q4)	13 (22.8)	643 (21.2)	2.36 (0.98, 5.67)	
		P_{trend}			0.67	0.73
Jass Type 2	Females	Lowest (Q1)	19 (18.3)	768 (20.2)	1 (ref)	

		Second (Q2)	21 (20.2)	1139 (29.9)	1.21 (0.64, 2.27)	
		Third (Q3)	34 (32.7)	1002 (26.3)	2.01 (1.10, 3.68)	
		Highest (Q4)	30 (28.8)	901 (23.6)	1.99 (1.07, 3.72)	
		P_{trend}			0.50	0.19
	Males	Lowest (Q1)	5 (12.8)	879 (29.1)	1 (ref)	
		Second (Q2)	11 (28.2)	775 (25.6)	2.19 (0.75, 6.43)	
		Third (Q3)	12 (30.8)	728 (24.1)	2.30 (0.78, 6.77)	
		Highest (Q4)	11 (28.2)	643 (21.2)	2.35 (0.78, 7.11)	
		P_{trend}			0.72	0.82
Jass type 3	Females	Lowest (Q1)	147 (20.9)	1139 (29.9)	1 (ref)	
		Second (Q2)	170 (24.1)	1002 (26.3)	1.34 (1.05, 1.70)	
		Third (Q3)	185 (26.3)	901 (23.6)	1.70 (1.32, 2.17)	
		Highest (Q4)	202 (28.7)	768 (20.2)	2.17 (1.69, 2.80)	
		P_{trend}			0.0087	0.044

	Males	Lowest (Q1)	104 (19.9)	879 (29.1)	1 (ref)	
		Second (Q2)	115 (22.0)	775 (25.6)	1.26 (0.95, 1.67)	
		Third (Q3)	150 (28.7)	728 (24.1)	1.83 (1.39, 2.41)	
		Highest (Q4)	152 (29.3)	643 (21.2)	2.26 (1.70, 3.01)	
		P_{trend}			8.45e-03	0.08
Jass type 4	Females	Lowest (Q1)	202 (19.3)	1139 (29.9)	1 (ref)	
		Second (Q2)	231 (22.8)	1002 (26.3)	1.42 (1.15, 1.75)	
		Third (Q3)	284 (27.2)	901 (23.6)	2.13 (1.71, 2.63)	
		Highest (Q4)	329 (31.4)	768 (20.2)	3.04 (2.44, 3.79)	
		P_{trend}			2.61e-06	(ref)
	Males	Lowest (Q1)	218 (18.6)	879 (29.1)	1 (ref)	
		Second (Q2)	246 (21.0)	775 (25.6)	1.27 (1.02, 1.57)	
		Third (Q3)	302 (25.7)	728 (24.1)	1.75 (1.42, 2.15)	
		Highest (Q4)	408 (34.8)	643 (21.2)	2.86 (2.32, 3.52)	

		P_{trend}			3.54e-06	(ref)
Jass type 5	Females	Lowest (Q1)	5 (7.2)	1139 (29.9)	1 (ref)	
		Second (Q2)	21 (30.4)	1002 (26.3)	5.03 (1.88, 13.50)	
		Third (Q3)	22 (31.9)	901 (23.6)	6.42 (2.35, 17.53)	
		Highest (Q4)	21 (30.4)	768 (20.2)	7.18 (2.57, 20.01)	
		P_{trend}			0.17	0.78
	Males	Lowest (Q1)	17 (28.8)	879 (29.1)	1 (ref)	
		Second (Q2)	13 (22.0)	775 (25.6)	0.71 (0.34, 1.48)	
		Third(Q3)	17 (28.8)	728 (24.1)	0.92 (0.45, 1.88)	
		Highest (Q4)	12 (20.3)	643 (21.2)	0.69 (0.31, 1.52)	
		P_{trend}			1	0.015

IV. Discussion

In this study, we evaluated an ERS as a composite measure of environmental risk factors for CRC, and sought to determine whether associations with that ERS would vary according to CRC molecular attributes. The results from the investigation into the individual tumor markers suggest that the ERS is positively associated with risk, regardless of *BRAF* or *KRAS* mutation, MSI, and CIMP status. Additionally, the results also suggest that environmental risk may be slightly more strongly associated with risk of *KRAS*-wildtype CRC than with *KRAS*-mutated

CRC, particularly in females. The ERS was also positively associated with CRC regardless of Jass molecular subtype. When stratified by sex, the results suggested that environmental risk may be slightly more strongly associated with the risk of Jass type 4 CRC when compared to Jass type 5 CRC in males, and in females, the results suggested that environmental risk may be slightly more strongly associated with the risk of Jass type 4 CRC compared to Jass type 3 CRC.

Associations of environmental risk factors with CRC risk have been documented in previous studies. Jeon et al. and Wang et al. found a significant association between an overall E-score that was very similar to the current one and CRC risk in both men and women (15,16). Archambault et al.'s ERS score was found to be significantly associated with early-onset CRC (14). The consistency is not surprising due to the partially overlapping studies.

Previous studies have demonstrated associations between elements of our ERS and CRC molecular attributes. High intake of fiber, fruits, vegetables, and calcium had inconsistent associations, with some studies finding a reduced risk of *KRAS*-mutated CRC (17,20-23) and others finding no difference in association by *KRAS* status (17,21,24–26). High intake of red meat and alcohol had inconsistent results in their associations as well, with some studies finding an increased risk of *KRAS* mutations in CRC cases related to high intake of red meat and alcohol (17,20,27) and others finding no difference in their associations with *KRAS* status when comparing high intake of alcohol and red meat to low intake of alcohol and red meat (2,17,20,21,26,28,29). High BMI, high levels of physical activity, PMH use, and high use of NSAIDs were found to produce no difference in their associations with CRC risk when stratified by *KRAS* status (21,24). Smoking was significantly associated with *KRAS*-mutated tumors when compared to *KRAS*-wildtype tumors in Wang et al, although this is not independent of our study as the same dataset was used in that study (12). This is inconsistent with our findings in that we observed modest differences in associations between the ERS and CRC risk according to *KRAS* mutation status, with a slightly stronger association observed on *KRAS* wild-type tumors. It is unclear which environmental factors are the most influential in their influence over *KRAS*-wildtype tumorigenesis.

Wang et al. found that associations between smoking habits and CRC differed when comparing Jass types 1, 2, and 5 to Jass type 4 (12). Murphy et al. found that all Jass types showed consistency in their association with increasing CRC risk and BMI in comparison to Jass type 4, though types 2 and 5 showed modestly stronger associations and higher BMI was associated with with an elevated risk of CRC with Jass types 2, 3, and 4 (30). This is somewhat inconsistent with our results, as we found a modestly stronger association between ERS and CRC risk in type 4 when compared to types 3 and 5, but not types 1 and 2.

Observed heterogeneity in risk factor associations according to individual tumor markers and molecular subtypes likely reflect differences in tumor etiology. There are three pathways for CRC tumorigenesis according to Legget and Whitehall: traditional, alternate, and serrated (31). Jass types 4 and 5 are associated with the traditional pathway (11,31). Jass type 3 is associated with the alternate pathway (31). Types 1 and 2 are associated with the serrated pathway (31). These different pathways result in divergent tumor biology, and have relevance to associations

observed here. Jass types 1, 2, 4 and 5 are associated with *KRAS*-wildtype tumors and type 3 is associated with *KRAS*-mutated tumors. We found that *KRAS*-wildtype tumors had a modestly higher association between the ERS and CRC than their *KRAS*-mutated counterparts and also found that tumors with Jass type 3 (the only Jass type that is characterized by a *KRAS* mutation) had a modestly lower association with the ERS and CRC than Jass type 4. This suggests that the role of environmental risk in *KRAS*-mutated tumors may be more subdued than in wildtype CRC and, furthermore, the role of environmental risk may thus be more subdued in the alternate pathway of tumorigenesis than the traditional pathway.

Since *BRAF* mutation, CIMP and MSI status were not found to have a differential association between the ERS and CRC, this suggests that the serrated pathway is not any more influenced by environmental risk than the traditional pathway. This is consistent with other findings in the literature (17,25,28,32–39). However, individual environmental factors may still be differentially associated with CRC risk according to different etiologic pathways.

Our results should be interpreted in the context of study limitations. First of all, many of the factors incorporated into the ERS were self-reported. As we used both cohort studies and case-control studies in our analysis, this can lead to both attenuation of the results and bias that differs by case-control status, such as recall bias. However, previous studies have shown that self-reported measures show a moderate to high amount of accuracy (40). Additionally, previous studies using both case-control and prospective studies from the GECCO consortium have shown consistency in the direction of associations, which indicates that these biases related to case-control studies are likely small (12,30). Secondly, this study only included non-Hispanic-White participants, as our numbers for other racial and ethnic groups were very small and therefore excluded. This poses an issue with generalization as our findings to other racial and ethnic populations. This is especially problematic given that CRC is shown to have higher rates in non-White populations such as African-American and Alaska Native adults. More studies must be done to examine environmental exposures in these populations and link them to tumor subtypes.

Despite these limitations, our study also has many strengths. To the best of our knowledge, no other study has assessed whether or not these tumor markers and molecular subtypes are differentially associated with an environmental risk score. Many studies have looked at whether or not these markers and subtypes are associated with individual risk factors (20,21) but none have looked at environmental risk as a whole. Furthermore, many studies have only analyzed individual tumor markers, but given our large sample size our study was able to examine Jass types. This allowed us to investigate many paths of tumorigenesis (e.g., traditional, alternate, and serrated pathways of carcinogenesis). Lastly, we used a standardized data harmonization approach which leads to better comparability between studies.

In conclusion, we found that *KRAS*-wildtype tumors had a slightly stronger association with environmental risk and CRC than their mutated counterparts. The association between the ERS and CRC was consistent and statistically significant across *BRAF* mutation status, CIMP status, and MSI status. Associations between environmental risk and CRC were uniform,

regardless of Jass type. More must be done to investigate these associations to determine the underlying biological mechanisms and more analyses should be conducted in relation to individual environmental risk factors and Jass types.

V. References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018 Nov;68(6):394–424.
2. O'Brien MJ, Yang S, Mack C, Xu H, Huang CS, Mulcahy E, et al. Comparison of Microsatellite Instability, CpG Island Methylation Phenotype, BRAF and KRAS Status in Serrated Polyps and Traditional Adenomas Indicates Separate Pathways to Distinct Colorectal Carcinoma End Points. Am J Surg Pathol. 2006 Dec;30(12):1491–501.
3. Keum N, Giovannucci E. Global burden of colorectal cancer: emerging trends, risk factors and prevention strategies. Nat Rev Gastroenterol Hepatol. 2019 Dec;16(12):713–32.
4. Baena R, Salinas P. Diet and colorectal cancer. Maturitas. 2015 Mar 1;80(3):258–64.
5. McNabb S, Harrison TA, Albanes D, Berndt SI, Brenner H, Caan BJ, et al. Meta-analysis of 16 studies of the association of alcohol with colorectal cancer. Int J Cancer. 2020 Feb 1;146(3):861–73.
6. Yusof AS, Isa ZM, Shah SA. Dietary Patterns and Risk of Colorectal Cancer: A Systematic Review of Cohort Studies (2000-2011). Asian Pac J Cancer Prev. 2012;13(9):4713–7.
7. Murphy N, Moreno V, Hughes DJ, Vodicka L, Vodicka P, Aglago EK, et al. Lifestyle and dietary environmental factors in colorectal cancer susceptibility. Mol Aspects Med. 2019 Oct 1;69:2–9.
8. Botteri E, Borroni E, Sloan EK, Bagnardi V, Bosetti C, Peveri G, et al. Smoking and Colorectal Cancer Risk, Overall and by Molecular Subtypes: A Meta-Analysis. Off J Am Coll Gastroenterol ACG. 2020 Dec;115(12):1940–9.
9. Wolin KY, Yan Y, Colditz GA, Lee IM. Physical activity and colon cancer prevention: a meta-analysis. Br J Cancer. 2009 Feb;100(4):611–6.
10. Green J, Czanner G, Reeves G, Watson J, Wise L, Roddam A, et al. Menopausal hormone therapy and risk of gastrointestinal cancer: Nested case–control study within a prospective cohort, and meta-analysis. Int J Cancer. 2012;130(10):2387–96.
11. Jass JR. Classification of colorectal cancer based on correlation of clinical, morphological and molecular features. Histopathology. 2007;50(1):113–30.
12. Wang X, Amitay E, Harrison TA, Banbury BL, Berndt SI, Brenner H, et al. Association Between Smoking and Molecular Subtypes of Colorectal Cancer. JNCI Cancer Spectr. 2021 Aug;5(4).
13. Phipps AI, Alwers E, Harrison T, Banbury B, Brenner H, Campbell PT, et al. Association Between Molecular Subtypes of Colorectal Tumors and Patient Survival, Based on Pooled Analysis of 7 International Studies. Gastroenterology. 2020 Jun;158(8):2158-2168.e4.
14. Archambault AN, Jeon J, Lin Y, Thomas M, Harrison TA, Bishop DT, et al. Risk Stratification for Early-Onset Colorectal Cancer Using a Combination of Genetic and Environmental Risk Scores: An International Multi-Center Study. JNCI J Natl Cancer Inst. 2022 Jan 13;114(4):528–39.
15. Jeon J, Du M, Schoen RE, Hoffmeister M, Newcomb PA, Berndt SI, et al. Determining Risk

of Colorectal Cancer and Starting Age of Screening Based on Lifestyle, Environmental, and Genetic Factors. Gastroenterology. 2018 Jun;154(8):2152-2164.e19.
16. Wang X, O'Connell K, Jeon J, Song M, Hunter D, Hoffmeister M, et al. Combined effect of modifiable and non-modifiable risk factors for colorectal cancer risk in a pooled analysis of 11 population-based studies. BMJ Open Gastroenterol. 2019 Dec 2;6(1):e000339.
17. Slattery ML, Curtin K, Sweeney C, Levin TR, Potter J, Wolff RK, et al. Diet and lifestyle factor associations with CpG island methylator phenotype and BRAF mutations in colon cancer. Int J Cancer. 2007;120(3):656–63.
18. Genetics and Epidemiology of Colorectal Cancer Consortium [Internet]. Fred Hutch. [cited 2022 Oct 19]. Available from: https://www.fredhutch.org/en/research/divisions/public-health-sciences-division/research/cancer-prevention/genetics-epidemiology-colorectal-cancer-consortium-gecco.html
19. Brenner H, Hoffmeister M, Arndt V, Haug U. Gender differences in colorectal cancer: implications for age at initiation of screening. Br J Cancer. 2007 Mar;96(5):828–31.
20. El Asri A, Zarrouq B, El Kinany K, Bouguenouch L, Ouldim K, El Rhazi K. Associations between nutritional factors and KRAS mutations in colorectal cancer: a systematic review. BMC Cancer. 2020 Jul 28;20(1):696.
21. Naguib A, Mitrou PN, Gay LJ, Cooke JC, Luben RN, Ball RY, et al. Dietary, lifestyle and clinicopathological factors associated with BRAF and K-ras mutations arising in distinct subsets of colorectal cancers in the EPIC Norfolk study. BMC Cancer. 2010 Mar 16;10(1):99.
22. Laso N, Mas S, Jose Lafuente M, Casterad X, Trias M, Ballesta A, et al. Decrease in specific micronutrient intake in colorectal cancer patients with tumors presenting Ki-ras mutation. Anticancer Res. 2004;24(3b):2011–20.
23. Bautista D, Obrador A, Moreno V, Cabeza E, Canet R, Benito E, et al. Ki-ras mutation modifies the protective effect of dietary monounsaturated fat and calcium on sporadic colorectal cancer. Cancer Epidemiol Biomarkers Prev. 1997 Jan 1;6(1):57–61.
24. Wark PA, Van der Kuil W, Ploemacher J, Van Muijen GNP, Mulder CJJ, Weijenberg MP, et al. Diet, lifestyle and risk of K-ras mutation-positive and -negative colorectal adenomas. Int J Cancer. 2006;119(2):398–405.
25. Keum N, Liu L, Hamada T, Qian ZR, Nowak JA, Cao Y, et al. Calcium intake and colon cancer risk subtypes by tumor molecular characteristics. Cancer Causes Control. 2019 Jun 1;30(6):637–49.
26. Martínez ME, Maltzman T, Marshall JR, Einspahr J, Reid ME, Sampliner R, et al. Risk factors for Ki-ras protooncogene mutation in sporadic colorectal adenomas. Cancer Res. 1999 Oct 15;59(20):5181–5.
27. Carr PR, Jansen L, Bienert S, Roth W, Herpel E, Kloor M, et al. Associations of red and processed meat intake with major molecular pathological features of colorectal cancer. Eur J Epidemiol. 2017 May 1;32(5):409–18.
28. Bongaerts BWC, Goeij AFPM de, Vogel S de, Brandt PA van den, Goldbohm RA, Weijenberg MP. Alcohol consumption and distinct molecular pathways to colorectal cancer. Br J Nutr. 2007 Mar;97(3):430–4.
29. Kampman E, Voskuil DW, van Kraats AA, Balder HF, van Muijen GN, Goldbohm RA, et al. Animal products and K-ras codon 12 and 13 mutations in colon carcinomas. Carcinogenesis. 2000 Feb;21(2):307–9.
30. Murphy N, Newton CC, Song M, Papadimitriou N, Hoffmeister M, Phipps AI, et al. Body

mass index and molecular subtypes of colorectal cancer. JNCI J Natl Cancer Inst. 2023 Feb 1;115(2):165–73.
31. Leggett B, Whitehall V. Role of the Serrated Pathway in Colorectal Cancer Pathogenesis. Gastroenterology. 2010 May 1;138(6):2088–100.
32. Limsui D, Vierkant RA, Tillmans LS, Wang AH, Weisenberger DJ, Laird PW, et al. Cigarette Smoking and Colorectal Cancer Risk by Molecularly Defined Subtypes. JNCI J Natl Cancer Inst. 2010 Jul 21;102(14):1012–22.
33. Nishihara R, Morikawa T, Kuchiba A, Lochhead P, Yamauchi M, Liao X, et al. A Prospective Study of Duration of Smoking Cessation and Colorectal Cancer Risk by Epigenetics-related Tumor Classification. Am J Epidemiol. 2013 Jul 1;178(1):84–100.
34. Hughes LAE, Simons CCJM, Brandt PA van den, Goldbohm RA, Goeij AF de, Bruïne AP de, et al. Body Size, Physical Activity and Risk of Colorectal Cancer with or without the CpG Island Methylator Phenotype (CIMP). PLOS ONE. 2011 Apr 5;6(4):e18571.
35. Razzak AA, Oxentenko AS, Vierkant RA, Tillmans LS, Wang AH, Weisenberger DJ, et al. Alcohol Intake and Colorectal Cancer Risk by Molecularly Defined Subtypes in a Prospective Study of Older Women. Cancer Prev Res (Phila Pa). 2011 Dec 4;4(12):2035–43.
36. Diergaarde B, Braam H, van Muijen GNP, Ligtenberg MJL, Kok FJ, Kampman E. Dietary Factors and Microsatellite Instability in Sporadic Colon Carcinomas. Cancer Epidemiol Biomarkers Prev. 2003 Dec 1;12(11):1130–6.
37. Hughes LAE, Simons CCJM, van den Brandt PA, van Engeland M, Weijenberg MP. Lifestyle, Diet, and Colorectal Cancer Risk According to (Epi)genetic Instability: Current Evidence and Future Directions of Molecular Pathological Epidemiology. Curr Colorectal Cancer Rep. 2017 Dec 1;13(6):455–69.
38. Brändstedt J, Wangefjord S, Borgquist S, Nodin B, Eberhard J, Manjer J, et al. Influence of anthropometric factors on tumour biological characteristics of colorectal cancer in men and women: a cohort study. J Transl Med. 2013 Nov 21;11(1):293.
39. Hoffmeister M, Bläker H, Kloor M, Roth W, Toth C, Herpel E, et al. Body Mass Index and Microsatellite Instability in Colorectal Cancer: A Population-based Study. Cancer Epidemiol Biomarkers Prev. 2013 Dec 5;22(12):2303–11.
40. Hu FB, Satija A, Rimm EB, Spiegelman D, Sampson L, Rosner B, et al. Diet Assessment Methods in the Nurses' Health Studies and Contribution to Evidence-Based Nutritional Policies and Guidelines. Am J Public Health. 2016 Sep;106(9):1567–72.