Cancer Causes & Control (2019) 30:1127–1135
https://doi.org/10.1007/s10552-019-01210-1

ORIGINAL PAPER

Substitution of dietary protein sources in relation to colorectal cancer risk in the NIH-AARP cohort study

Linda M. Liao[1,2] · Erikka Loftfield[1] · Arash Etemadi[1] · Barry I. Graubard[1] · Rashmi Sinha[1]

Received: 26 March 2019 / Accepted: 11 July 2019 / Published online: 20 July 2019

Abstract

Purpose To evaluate the substitution effect of plant for animal protein with risk of CRC in the large prospective National Institutes of Health-AARP cohort study.

Methods Protein intake was assessed at baseline using a food frequency questionnaire. HRs and 95% CIs were estimated using multivariable adjusted hazard ratios from Cox proportional hazards models. We used a substitution model with total protein intake held constant, so that an increase in plant protein was offset by an equal decrease in animal protein.

Results Among 489,625 individuals, we identified 8,995 incident CRCs after a median follow-up of 15.5 years. Substituting plant protein for animal protein was associated with a reduced risk of CRC (HR for highest vs. lowest fifth 0.91; 95% CI 0.83–0.99). This reduction in CRC risk appeared to be primarily due to substituting plant protein for red meat protein (HR 0.89; 95% CI 0.81–0.97), not white meat protein (HR 0.96; 95% CI 0.88–1.05) or other animal protein (HR 0.94; 95% CI 0.86–1.03). When further evaluated by source, reduction in CRC risk was limited to the substitution of protein from bread, cereal, and pasta for red meat protein (HR 0.86; 95% CI 0.80–0.93); this association was stronger for distal colon (HR 0.78; 95% CI 0.67–0.90) and rectal cancer (HR 0.79; 95% CI 0.68–0.91) but null for proximal colon (HR 0.99; 95% CI 0.88–1.11).

Conclusions This study shows that substituting plant protein for animal protein, especially red meat protein, is associated with a reduced risk of CRC, and suggests that protein source impacts CRC risk.

Keywords Dietary protein · Protein source · Plant protein · Colorectal cancer · Prospective cohort study

Introduction

The relationship between red and processed meat consumption and colorectal cancer risk continues to be debated despite a recent report by the International Agency for Cancer that classified red meat as "probably carcinogenic" and processed meat as "carcinogenic" to humans [1]. Previous evaluations of meat subtypes and cancer risk in prospective studies, including the NIH-AARP Diet and Health Study, have demonstrated positive associations between red meat intake and several cancers [2–4], including colorectal cancer [5–7]. Exposure to potentially carcinogenic compounds, such as polycyclic aromatic hydrocarbons or heterocyclic amines, nitrate/nitrite and N-nitroso compounds, heme iron, and animal fat [1], all of which are largely absent from plant-based foods, may explain the observed associations between animal protein and cancer. However, little is known about the impact of substituting plant for animal protein in relation to colorectal cancer risk [8, 9].

The most recent meta-analysis of observational studies did not find an association between dietary protein intake and colorectal cancer risk, but the authors acknowledged that a high degree of study heterogeneity, most likely due to population differences in dietary intake, impacted their results [9]. Past studies that evaluated protein intake and cancer risk did not always discern between animal and plant protein, and few studies have directly examined whether substituting one protein source for another has an impact on health. Epidemiologic evidence indicates that protein source, animal or plant, alters chronic disease and mortality associations, but the biological mechanism remains unclear [8, 10, 11]. Large

✉ Linda M. Liao
linda.liao@nih.gov

[1] Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA

[2] Division of Cancer Epidemiology and Genetics, National Cancer Institute, 9609 Medical Center Dr., 6E424, Rockville, MD 20850, USA

prospective investigations of the relationship between protein source and colorectal cancer are lacking. Therefore, we evaluated the effect of substituting plant protein for animal protein on risk of colorectal cancer overall and by subsite among nearly a half-million U.S. adults who were followed for more than a decade. Finally, we evaluated the effect of substituting animal protein for plant protein on associations with colorectal cancer risk.

Subjects and methods

Study population

The NIH-AARP Diet and Health study is a large prospective cohort of U.S. men and women, aged 50 to 71 years who resided in one of the six U.S. states (California, Florida, Louisiana, New Jersey, North Carolina, and Pennsylvania) or two U.S. metropolitan areas (Atlanta, Georgia and Detroit, Michigan) at baseline. Details of the NIH-AARP study have been described previously [12]. Briefly, in 1995–1996, a self-administered questionnaire was mailed to 3.5 million AARP members. The baseline questionnaire queried about demographics, lifestyle factors, and dietary intake. In total, 566,398 participants returned the questionnaires with satisfactory dietary data and constituted the baseline cohort. A second questionnaire was sent out 6 months later to respondents to collect more detailed information on specific risk factors for cancer (RFQ), including non-steroidal anti-inflammatory drugs (NSAID) use; 334,906 participants completed the second questionnaire. The study was approved by the Special Studies Institutional Review Board of the National Cancer Institute.

Among the 566,398 cohort participants who satisfactorily completed the baseline questionnaire and gave written informed consent by virtue of completing and returning the questionnaire, we excluded proxy-responders ($n = 15{,}760$), individuals with prevalent cancer except non-melanoma skin cancer ($n = 51{,}334$) or self-reported end-stage renal disease at baseline ($n = 997$), and individuals with only a death record for cancer ($n = 4{,}253$). Furthermore, we excluded individuals who reported extreme intakes (> 2 times the interquartile ranges of Box–Cox log-transformed intake) of total energy ($n = 4{,}372$) and individuals with 0 years of follow-up ($n = 57$). After exclusions, the analytic cohort was comprised of 489,625 individuals (291,794 men and 197,831 women).

Dietary intake assessment

Dietary intake was assessed at baseline using a self-administered 124-item food frequency questionnaire (FFQ), the NCI-Diet History Questionnaire (DHQ). The FFQ asked about usual frequency of intake and portion size over the past 12 months using ten predefined frequency categories ranging from never to 6 + times per day for beverages and from never to 2 + times per day for solid foods as well as three categories of portion size. The food items, portion sizes, and nutrient database were constructed using the U.S. Department of Agriculture's 1994–1996 Continuing Survey of Food Intakes by Individuals. The FFQ was validated for foods and nutrients using two non-consecutive 24-h dietary recalls within a year of baseline in a subset of participants [13].

Total dietary protein intake was categorized into animal protein and plant protein. Animal protein was further categorized into different sources: red meat, white meat, and other animal protein (e.g., dairy products and eggs). Red meat included all types of fresh (beef, pork, hamburger, steak, liver, and meats in foods such as chili, lasagna, and stew) and processed red meat (bacon, cold cuts, ham, hot dogs, and sausage). White meat included chicken, turkey, and fish, canned tuna, and processed white meat (poultry cold cuts, low-fat sausages, and low-fat hotdogs made from poultry). Plant protein was further categorized into different sources: bread, cereal and pasta; nuts; beans and legumes; and other plant protein (e.g., fruits, vegetables).

Cohort follow-up and cancer ascertainment

Incident cancer cases were identified through probabilistic linkage with state cancer registries in the eight original states and three additional states (Arizona, Nevada, and Texas) where participants most commonly moved during follow-up. A previous validation study estimated that the cancer registry linkage with eight states identified about 90% of all incident cancers in this cohort [14]. Linkage to 11 state registries has increased our likelihood of capturing cases, but we are unable to capture cases that were diagnosed outside of these states or were not captured by a cancer registry due to various reasons. Follow-up began from enrollment in 1995–1996 and continued until cancer diagnosis, movement out of the cancer registry area, loss to follow-up, death, or 31 December 2011, whichever came first. Vital status was obtained by linkage to the National Death Index and cancer registry linkage. Colorectal cancers were defined by anatomic sites and histology codes as defined by the third edition of the International Classification of Diseases for Oncology and included codes C180–C189, C199, C209, and C260. Colorectal subsites were further classified as proximal colon (C180–C184), distal colon (C185–C187), and rectum (C199 and C209). We restricted our analysis to adenocarcinomas, excluding lymphomas, sarcomas, neuroendocrine tumors, squamous cell tumors, other non-adenocarcinoma histology types, and cases with unspecified histologies.

Statistical analyses

All dietary variables were adjusted for total energy intake using the residual method [15]. Cox proportional hazards regression models with person-years as the time metric was used to estimate hazard ratios (HRs) and 95% confidence intervals (95% CIs) for colorectal cancer in relation to quintiles of dietary protein intake. Proportional hazards assumption was tested by including an interaction term between a time-dependent binary covariate, which indicated if follow-up was in the first 7.5 years or the second 7.5 years, and the quintile terms for protein intake. No deviation from the proportional hazards assumption was detected. We also examined the potential non-linearity of the association between plant protein intake and colorectal cancer risk by using a cubic regression spline and concluded that the association was non-linear ($p = 0.002$).

We used a substitution model with adjustment for total protein intake, such that an increase in plant protein is offset by an equal decrease in animal protein and total protein intake remains constant [16]. Similarly, an increase in animal protein is offset by an equal decrease in plant protein and total protein intake remains constant. For substitution models evaluating the substitution of plant protein for different sources of animal protein, two of the three sources of animal protein intake were simultaneously included in the model with total protein intake. In this model, an increase in plant protein is offset by an equal decrease in the source of animal protein not included in the model and total protein intake remains constant. We analyzed the effects of substituting different sources of plant protein for red meat. In these analyses, we used substitution models where each component of plant protein intake, white meat protein, other animal protein, and total protein intake were simultaneously included in the model. In this model, an increase in each respective source of plant protein is offset by an equal decrease in red meat protein where total protein intake remains constant. A similar substitution model was used to evaluate the components of animal protein intake, in which each effect represents a substitution of plant protein.

In addition to adjustment for total protein intake, we adjusted for potential dietary and lifestyle confounding factors in the multivariable analyses, including age, total energy, sex, education, marriage status, family history of colon cancer, race, body mass index (BMI), smoking, frequency of vigorous physical activity, alcohol intake, fruit intake, vegetable intake, total calcium intake (diet + supplemental), total folate intake (diet + supplemental), and dietary fiber intake. We also conducted several secondary analyses including additional adjustment for NSAID use among those who completed a follow-up questionnaire, evaluation of a potential interaction with gender, and exclusion of outcomes that occurred in the first 2 years of follow-up. Tests for linear trend across quintiles of protein intake were performed using the median values of each exposure category as a continuous variable in the model. Statistical significance was defined as $p < 0.05$, and all tests were 2-sided. Statistical analyses were performed using SAS 9.4 (SAS Institute, Inc.).

Results

During a median follow-up of 15.5 years, we identified 8,995 incident colorectal cancers (6,719 colon and 2,276 rectum). Of the colon cancer cases, 3,990 were proximal colon cancers, 2,514 were distal colon cancers, and 215 had an unknown subsite location. Compared with individuals in the lowest quintile of animal protein intake, individuals in the highest quintile of animal protein intake were more likely to be White non-Hispanic, have slightly higher BMIs and less physically active, in addition to consuming more animal products and less fruits (Table 1). Individuals in the highest quintile of plant protein intake were more likely to be college-educated, physically active, and consume more fruits and vegetables, and were less likely to smoke and consume red meat compared with individuals in the lowest quintile of plant protein intake. On average, animal protein accounted for 60% and plant protein accounted for 40% of total protein consumption. The largest contributors to total protein intake (means) was "Other animal protein" 22.6%; "White meat protein" 18.8%; "Protein from bread, cereal, and pasta" 18.4%; and "Red meat protein" 18.4%. A large proportion of plant protein was from bread, cereal, and pasta (mean: 46%; interquartile range, IQR 38–54%). This is consistent with analyses of data from NHANES that have found that plant protein represented one-third of total dietary protein intake, with grains being the primary source of plant protein intake [17].

A higher amount of animal protein intake was positively associated with a higher risk of colorectal cancer overall (HR 1.15; 95% CI 1.03–1.28; p trend = 0.005) but was not statistically significant for all subsites (Fig. 1). In Table 2, we present the risk of colorectal cancer overall and by cancer subsite associated with the substitution of plant protein for animal protein and vice versa, while simultaneously controlling for total protein. In general, the substitution of plant protein for animal protein was inversely associated with risk of colorectal cancer overall and for rectal and distal cancer. The HR for the highest category of plant protein substitution was 0.91 (95% CI 0.83–0.99; p trend = 0.03) for colorectal cancer risk, 0.84 (95% CI 0.71–1.00) for rectal cancer risk, and 0.83 (95% CI 0.71–0.98) for distal cancer risk. No association was observed for proximal colon cancer (Table 2).

When we further evaluated the substitution of plant protein for animal protein by animal protein source, the reduction in colorectal cancer risk appeared to be primarily due

Table 1 Distribution of baseline characteristics of the NIH-AARP diet and health study cohort ($n = 489{,}625$)

Characteristic	Type of protein									
	Animal					Plant				
	Q1	Q2	Q3	Q4	Q5	Q1	Q2	Q3	Q4	Q5
Age, mean (years)	62.1	62.3	62.2	62	61.5	61.6	62.1	62.2	62.2	62
Male sex (%)	63.5	55.7	55.6	57.6	65.6	66.4	54.5	54.3	57.8	65.3
BMI, mean (kg/m^2)	26.3	26.6	27	27.4	28.2	27.5	27.2	27.1	27	26.8
White, non-Hispanic (%)	87.7	90.5	92.1	93.2	92.8	90.1	90.9	92.1	92.3	90.7
College and postcollege (%)	37.2	37.8	38.3	39.6	40	32.5	35.8	38.8	41.3	44.5
Married or living as married (%)	67.5	66.6	67.8	69.5	71.6	68.1	66.2	67.5	69.8	71.2
Current smoker (%)	12.6	11.4	11.4	11.6	12.3	19.9	12.8	10.3	8.7	7.5
Physical activity (vigorous > 20 min) (%)										
Never	5	4.4	4.2	4	4.4	6.5	4.9	3.9	3.4	3.3
5 + times per week	23	19	17.4	17.7	19	15.7	15.8	17.7	20.5	26.3
Dietary intakes per day (mean[a])										
Total protein (g)	51.4	63.3	68.7	75.2	90.7	64.7	69.3	70.3	71.4	73.6
Animal protein (g)	22.1	35.5	41.8	48.8	66.2	47	45.6	43.8	41.8	36.3
Plant protein (g)	29.2	27.7	26.8	26.2	24.3	17.6	23.6	26.3	29.4	37.2
Red meat (g)	34.1	58	66.8	74.5	94.7	75.2	72.9	68.4	62.9	48.9
White meat (g)	28.9	46.8	56.2	68.1	98.9	54	60.3	61.6	62.4	60.5
Alcohol intake (g)	26.1	10.2	8.4	8.2	8.6	30.7	9.6	7.4	6.9	6.8
Fruit, MPED servings	2.6	2.1	2	1.9	1.7	1.6	1.9	2.1	2.2	2.4
Vegetables, MPED servings	2	2	1.9	2	2	1.3	1.7	1.9	2.1	2.8

[a]Residual energy adjustment values presented

to the substitution of plant protein for red meat protein (Table 2). In these models, two of the three components of animal protein were included simultaneously in the model with total protein intake; thus the component of animal protein not included in the model represented a substitution of plant protein for that component. Individuals in the highest compared to the lowest category of plant protein substitution for red meat protein had an 11 percent decreased risk of colorectal cancer (HR 0.89; 95% CI 0.81–0.97; p trend = 0.009). Similarly, the highest compared to the lowest category of plant protein substitution for red meat protein had a 16 percent decreased risk of rectal cancer (HR 0.84; 95% CI 0.71–1.00) and a 21 percent decreased risk of distal colon cancer (HR 0.79; 95% CI 0.67–0.93), but no association was observed for proximal colon cancer (HR 1.01; 95% CI 0.88–1.16). This reduction in colorectal cancer risk was not apparent with the substitution of plant protein for white meat protein or other animal protein (Table 2). Correspondingly, the substitution of red meat protein for plant protein was associated with a 15%, 23%, and 55%, statistically significant increased risk of proximal colon, distal colon, and rectal cancer, respectively (data not shown).

In Table 3, we examined the association of plant protein substitution with colorectal cancer by plant protein source. In these models, we used substitution models where each component of plant protein intake, white meat protein, other animal protein, and total protein intake were simultaneously included; thus HR estimates represent the substitution of a given source of plant protein for red meat protein, which was not included in the model. The reduction in colorectal cancer risk was limited to the substitution of protein from bread, cereal, and pasta for red meat protein (Table 3). Substituting protein from bread, cereal, and pasta for protein from red meat was associated with up to a 14% lower risk of colorectal cancer (95% CI 0.80–0.93; p trend < 0.0001). This association was stronger for distal colon (HR 0.78; 95% CI 0.67–0.90) and rectal cancer (HR 0.79; 95% CI 0.68–0.91) but was not observed for proximal colon cancer (HR 0.99; 95% CI 0.88–1.11). Substituting protein from nuts, beans, and legumes, or other plant protein for protein from red meat was not associated with colorectal cancer risk.

In secondary analyses, we further adjusted for NSAID use among those who completed a follow-up questionnaire ($n = 301{,}180$); however, observed associations did not change substantially (data not shown). We found no evidence of effect modification by gender, and associations were similar when we excluded outcomes that occurred in the first 2 years of follow-up (data not shown).

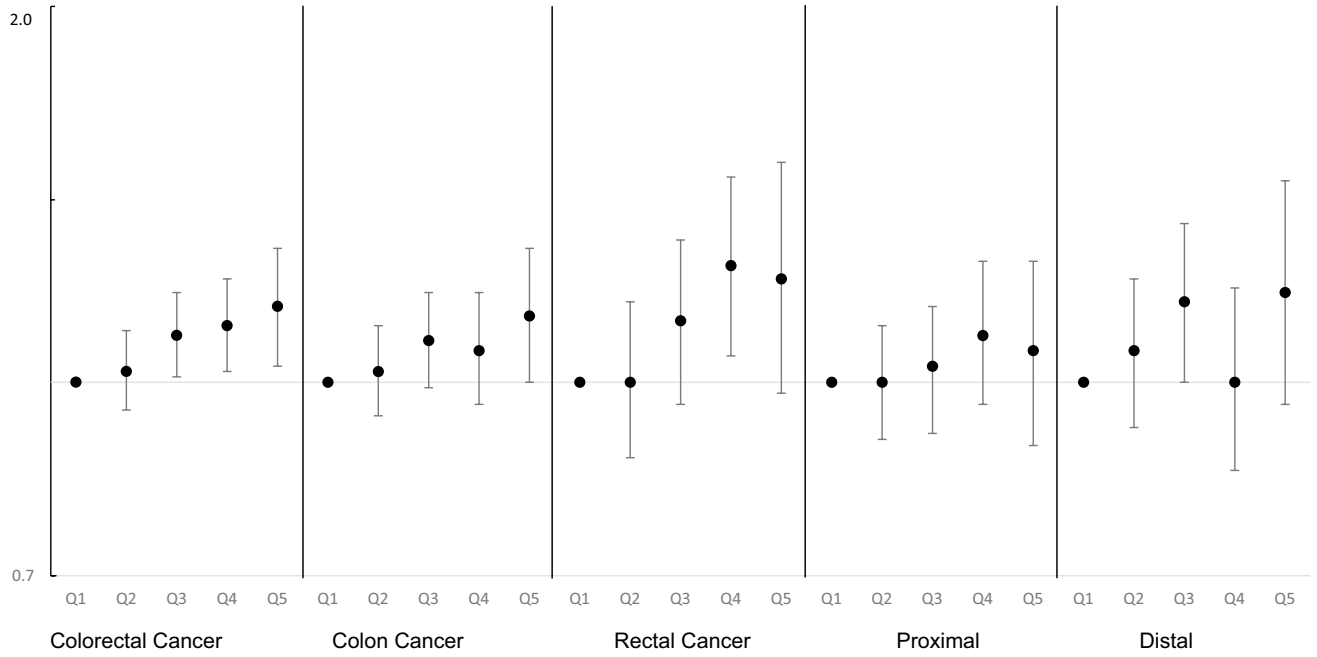

Fig. 1 Risk for colorectal cancer associated with substitution of animal protein for plant protein in the NIH-AARP diet and health study. We used a substitution model with adjustment for total protein intake, such that an increase in animal protein is offset by an equal decrease in plant protein and total protein intake remains constant. Adjusted for total protein intake (continuous), age (continuous), total energy, sex, education (less than high school, high school graduate or some college, and college graduate), marriage status (no, yes), family history of colon cancer (no, yes), race (non-Hispanic white, non-Hispanic black, and other), body mass index (BMI; 18.5 to < 25, 25 to < 30, 30 to < 35, ≥ 35 kg/m^2, unknown), smoking status (never, quit ≥ 10 years ago, quit 1–9 years ago, current smoker, or quit < 1 year ago, unknown), frequency of vigorous physical activity (never or rarely, 1–3 times/month, 1–2 times per week, 3–4 times per week, 5 or more times per week, or unknown), alcohol intake (none, > 0–5, ≥ 5–15, ≥ 15–30, ≥ 30 g per day), fruit intake (quintiles), vegetable intake (quintiles), total calcium intake (diet + supplemental), total folate intake (diet + supplemental), and dietary fiber intake

Discussion

In this large prospective investigation of protein intake in relation to colorectal cancer risk, we found that the substitution of plant protein for animal protein, especially red meat protein is associated with lower colorectal cancer risk. The substitution of plant protein for red meat protein was associated with up to an 11% decreased risk of colorectal cancer, with similar decreases observed among distal colon and rectal cancer. When further evaluated by protein sources, a statistically significant inverse association with colorectal cancer risk was limited to the substitution of protein from bread, cereal, and pasta for red meat protein, suggesting that protein source has an impact on risk of colorectal cancer. This inverse association was stronger for distal colon and rectal cancer but was not observed for proximal colon cancer.

A small number of epidemiologic studies have evaluated the association between dietary protein and colorectal cancer risk with conflicting results. However, few distinguished between protein type, and those were often limited by small case numbers. A recent meta-analysis of protein intake and colorectal cancer risk found no association; however, significant between study heterogeneity was evident in their results ($I^2 = 53.4\%$, $p_{het} = 0.0002$) [9]. Although the authors observed a non-significant inverse association for vegetable protein (RR = 0.85[0.60–1.20]; $n = 4$ studies), the subgroup analysis by protein type (animal vs. vegetable) was the only subgroup analysis that did not have high heterogeneity. Based on the subgroup results, the authors suggest that protein type was the potential contributor to the heterogeneity in their overall protein and colorectal cancer result. Furthermore, only one prospective cohort study included in this meta-analysis evaluated vegetable protein, and it was limited by a small number of cases ($n = 185$ cases) [18]. The remaining three studies were case–control studies. More recently, a few prospective studies have evaluated protein type in relation to chronic diseases and mortality, and plant protein intake has been associated with lower risk of cardiovascular disease, type 2 diabetes, and mortality [11, 19, 20]. Studies of plant protein and colorectal cancer risk are, however, lacking [9].

Our analysis also suggested that there are substantial differences in the association between plant protein intake and colorectal cancer by subsite. Several epidemiologic studies have suggested that risk factors differ in their

Table 2 Risk of colorectal cancer associated with substitution of plant protein for various animal protein sources

Animal protein source	Quintiles of plant protein intake					*p* trend
HR (95% CI)[a]	1	2	3	4	5	
Colorectal cancer						
All animal protein	1.00	0.90 (0.84–0.97)	0.87 (0.81–0.93)	0.88 (0.82–0.95)	0.91 (0.83–0.99)	0.04
Red meat protein	1.00	0.90 (0.84–0.97)	0.86 (0.80–0.93)	0.87 (0.81–0.94)	0.89 (0.81–0.97)	0.01
White meat protein	1.00	0.91 (0.85–0.97)	0.88 (0.82–0.95)	0.91 (0.84–0.98)	0.96 (0.88–1.05)	0.44
Other animal protein	1.00	0.90 (0.84–0.97)	0.87 (0.81–0.94)	0.90 (0.83–0.97)	0.94 (0.86–1.03)	0.22
Colon cancer						
All animal protein	1.00	0.93 (0.86–1.01)	0.88 (0.81–0.96)	0.91 (0.83–0.99)	0.94 (0.85–1.04)	0.20
Red meat protein	1.00	0.93 (0.86–1.00)	0.87 (0.80–0.95)	0.89 (0.82–0.98)	0.90 (0.81–1.00)	0.05
White meat protein	1.00	0.93 (0.86–1.01)	0.89 (0.82–0.97)	0.93 (0.85–1.01)	0.97 (0.87–1.08)	0.61
Other animal protein	1.00	0.93 (0.86–1.01)	0.89 (0.82–0.97)	0.93 (0.85–1.01)	0.97 (0.87–1.07)	0.54
Rectal cancer						
All animal protein	1.00	0.83 (0.72–0.95)	0.83 (0.72–0.96)	0.81 (0.70–0.94)	0.84 (0.71–1.00)	0.06
Red meat protein	1.00	0.83 (0.72–0.95)	0.84 (0.72–0.97)	0.81 (0.70–0.95)	0.84 (0.71–1.00)	0.07
White meat protein	1.00	0.84 (0.73–0.96)	0.86 (0.74–0.99)	0.85 (0.73–1.00)	0.93 (0.78–1.12)	0.52
Other animal protein	1.00	0.82 (0.72–0.94)	0.83 (0.72–0.96)	0.82 (0.71–0.96)	0.88 (0.74–1.05)	0.18
Proximal						
All animal protein	1.00	1.02 (0.91–1.13)	0.93 (0.83–1.04)	1.05 (0.93–1.17)	1.04 (0.91–1.19)	0.48
Red meat protein	1.00	1.01 (0.91–1.12)	0.92 (0.82–1.03)	1.04 (0.92–1.16)	1.02 (0.89–1.17)	0.71
White meat protein	1.00	1.02 (0.92–1.13)	0.94 (0.84–1.05)	1.07 (0.95–1.20)	1.08 (0.94–1.24)	0.22
Other animal protein	1.00	1.02 (0.92–1.13)	0.94 (0.84–1.05)	1.06 (0.95–1.19)	1.06 (0.93–1.22)	0.30
Distal						
All animal protein	1.00	0.84 (0.74–0.95)	0.83 (0.72–0.95)	0.74 (0.64–0.86)	0.84 (0.71–0.99)	0.02
Red meat protein	1.00	0.83 (0.73–0.94)	0.81 (0.71–0.93)	0.72 (0.62–0.84)	0.79 (0.67–0.94)	0.003
White meat protein	1.00	0.83 (0.73–0.95)	0.83 (0.72–0.95)	0.75 (0.65–0.87)	0.86 (0.72–1.01)	0.04
Other animal protein	1.00	0.84 (0.74–0.95)	0.83 (0.73–0.95)	0.76 (0.66–0.88)	0.87 (0.74–1.03)	0.06

[a]In order to evaluate the substitution of plant protein for each component of animal protein: plant protein, total protein, and the other two components of animal protein were included simultaneously in the model. The animal protein source that was not included in the model represents the animal protein source that plant protein was substituting for. Also adjusted for age (continuous), total energy, sex, education (less than high school, high school graduate or some college, and college graduate), marriage status (no, yes), family history of colon cancer (no, yes), race (non-Hispanic white, non-Hispanic black, and other), body mass index (BMI; 18.5 to < 25, 25 to < 30, 30 to < 35, ≥ 35 kg/m^2, unknown), smoking status (never, quit ≥ 10 years ago, quit 1–9 years ago, current smoker or quit < 1 year ago, unknown), frequency of vigorous physical activity (never or rarely, 1–3 times/month, 1–2 times per week, 3–4 times per week, 5 or more times per week, or unknown), alcohol intake (none, > 0–5, ≥ 5–15, ≥ 15–30, ≥ 30 g per day), fruit intake (quintiles), vegetable intake (quintiles), total calcium intake (diet + supplemental), total folate intake (diet + supplemental), and dietary fiber intake

association with tumor development by anatomic subsite. The proximal and distal portions of the large intestine have molecular and functional differences which results in sites being more susceptible to specific exposures, such as diet [21, 22]. In a recent pooled analysis of three prospective cohort studies, the authors observed that anatomical subsite modified the association between meat and risk of colorectal cancer with a distinct "right-to-left" trend of increasing risk as you progress from proximal colon to distal colon and rectum [7]. Similarly, our observed association with the substitution of plant protein was distinctly different between the proximal colon and more distal sites (distal colon and rectum) of the large intestine. Evaluating this association by subsite has not been thoroughly examined in the past, but our study suggests that it should be.

Our findings show that the association between dietary protein intake and colorectal cancer may differ by protein source, but the biological mechanism underlying this association is unclear. Animal protein is typically the main source of dietary protein in the United States and Europe and was also the main contributor to protein intake in our study [17, 23]. Higher protein intake corresponds with an increasing amount of undigested protein entering the colon [24]. This suggests that a large proportion of the undigested protein entering the colon is determined by the amount of animal protein intake, despite animal protein being slightly more digestible than plant protein. Fermentation by colonic

Table 3 Risk of colorectal cancer associated with substitution of different sources of plant protein for red meat protein

HR (95% CI)[a]	Quintiles of protein from bread, cereal, and pasta					*p* trend
	1	2	3	4	5	
Colorectal cancer	1.00	0.97 (0.91–1.03)	0.92 (0.86–0.98)	0.87 (0.81–0.94)	0.86 (0.80–0.93)	< 0.0001
Colon cancer	1.00	0.96 (0.89–1.04)	0.93 (0.86–1.01)	0.89 (0.82–0.96)	0.89 (0.82–0.97)	0.003
Rectal cancer	1.00	0.98 (0.86–1.11)	0.87 (0.76–1.00)	0.84 (0.73–0.96)	0.78 (0.67–0.91)	0.0004
Proximal	1.00	0.99 (0.89–1.09)	0.99 (0.89–1.10)	0.93 (0.84–1.04)	0.99 (0.88–1.11)	0.62
Distal	1.00	0.93 (0.83–1.06)	0.87 (0.76–0.99)	0.81 (0.71–0.92)	0.78 (0.67–0.90)	0.0001
	Quintiles of protein from nuts					
Colorectal cancer	1.00	0.96 (0.89–1.03)	0.96 (0.89–1.04)	0.99 (0.91–1.06)	0.95 (0.89–1.02)	0.26
Colon cancer	1.00	0.92 (0.85–1.00)	0.96 (0.88–1.05)	0.95 (0.87–1.04)	0.94 (0.87–1.02)	0.33
Rectal cancer	1.00	1.08 (0.94–1.24)	0.95 (0.81–1.10)	1.09 (0.94–1.27)	0.98 (0.86–1.12)	0.59
Proximal	1.00	0.90 (0.81–1.00)	0.93 (0.83–1.04)	0.93 (0.83–1.05)	0.93 (0.84–1.03)	0.53
Distal	1.00	0.92 (0.80–1.05)	0.98 (0.86–1.13)	0.93 (0.80–1.08)	0.90 (0.79–1.03)	0.16
	Quintiles of protein from beans and legumes					
Colorectal cancer	1.00	1.02 (0.95–1.09)	0.98 (0.91–1.06)	1.01 (0.94–1.08)	1.02 (0.94–1.10)	0.69
Colon cancer	1.00	1.00 (0.92–1.08)	0.96 (0.88–1.05)	0.99 (0.91–1.08)	0.99 (0.91–1.08)	0.94
Rectal cancer	1.00	1.07 (0.93–1.22)	1.04 (0.90–1.20)	1.04 (0.90–1.21)	1.09 (0.94–1.26)	0.37
Proximal	1.00	0.97 (0.88–1.08)	0.96 (0.86–1.08)	1.03 (0.92–1.15)	1.02 (0.91–1.14)	0.44
Distal	1.00	1.03 (0.91–1.18)	0.98 (0.85–1.12)	0.96 (0.84–1.10)	0.97 (0.84–1.12)	0.52
	Quintiles of other plant protein					
Colorectal cancer	1.00	1.02 (0.95–1.09)	0.97 (0.90–1.04)	0.98 (0.90–1.06)	1.01 (0.92–1.11)	0.94
Colon cancer	1.00	1.03 (0.95–1.11)	0.97 (0.88–1.06)	0.98 (0.89–1.08)	1.01 (0.91–1.13)	0.96
Rectal cancer	1.00	0.98 (0.85–1.13)	0.97 (0.84–1.13)	0.97 (0.82–1.14)	1.00 (0.83–1.20)	0.99
Proximal	1.00	1.06 (0.96–1.18)	0.95 (0.85–1.07)	1.00 (0.89–1.14)	1.06 (0.92–1.22)	0.61
Distal	1.00	0.96 (0.84–1.09)	0.99 (0.86–1.14)	0.94 (0.80–1.09)	0.94 (0.79–1.12)	0.49

[a]In order to evaluate the substitution of each component of plant protein for red meat protein: protein from bread/cereal/pasta, protein from nuts, protein from beans/legumes, other plant protein, white meat protein, other animal protein, and total protein were included simultaneously in order to evaluate substitution for red meat protein. Also adjusted for age (continuous), total energy, sex, education (less than high school, high school graduate or some college, and college graduate), marriage status (no, yes), family history of colon cancer (no, yes), race (non-Hispanic white, non-Hispanic black, and other), body mass index (BMI; 18.5 to < 25, 25 to < 30, 30 to < 35, ≥ 35 kg/m^2, unknown), smoking status (never, quit ≥ 10 years ago, quit 1–9 years ago, current smoker or quit < 1 year ago, unknown), frequency of vigorous physical activity (never or rarely, 1–3 times/month, 1–2 times per week, 3–4 times per week, 5 or more times per week, or unknown), alcohol intake (none, > 0–5, ≥ 5–15, ≥ 15–30, ≥ 30 g per day), fruit intake (quintiles), vegetable intake (quintiles), total calcium intake (diet + supplemental), total folate intake (diet + supplemental), and dietary fiber intake

bacteria of undigested protein into byproducts that have direct contact with the colonic mucosa is one of the potential mechanisms through which protein may play a role in the development of CRC [25]. These byproducts of protein fermentation, including ammonia, phenolic and indolic compounds, and hydrogen sulfide, may promote DNA damage and alterations in colonic epithelial cells [26]. As our results were primarily driven by the replacement of red meat protein, another explanation could be that the pathway through which this association occurs is driven by lower exposure to potentially carcinogenic compounds associated with red meat, such as heterocyclic amines, nitrate/nitrite and N-nitroso compounds, and heme iron [1].

This study is the first prospective cohort study to examine the substitution effect of plant protein for animal protein in relation to colorectal cancer risk. In addition to the advantages of using a cohort design, such as mitigating selection and recall bias, the main strength of our study was the large size, and relatively long follow-up which allowed us to conduct both subsite and subgroup analyses. Forty percent of our dietary protein was plant protein, a proportion similar to that of the American population [17]; this suggests that our cohort is representative of the US population with respect to the prevalence of plant protein intake. Measurement error in dietary assessments using an FFQ is well described in nutritional epidemiology, and non-differential misclassification resulting from this would likely result in an underestimation of the associations between protein intake and colorectal cancer risk. Our study is also limited by a single baseline dietary assessment; therefore, dietary changes over follow-up could not be accounted for. The FFQ did not assess soy

intake; however, soy intake accounts for a small percentage of protein intake in the United States according to data from NHANES [27]. It is also possible that although we controlled for many potential confounders, residual confounding from unmeasured or poorly measured factors may contribute to the observed association.

In summary, this study shows that the substitution of plant protein for animal protein, especially red meat protein, is associated with a lower risk of colorectal cancer, indicating that protein source has an impact on colorectal cancer risk and should be considered in future studies.

Acknowledgments This research was supported [in part] by the Intramural Research Program of the NIH, National Cancer Institute. Cancer incidence data from the Atlanta metropolitan area were collected by the Georgia Center for Cancer Statistics, Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, Georgia. Cancer incidence data from California were collected by the California Cancer Registry, California Department of Public Health's Cancer Surveillance and Research Branch, Sacramento, California. Cancer incidence data from the Detroit metropolitan area were collected by the Michigan Cancer Surveillance Program, Community Health Administration, Lansing, Michigan. The Florida cancer incidence data used in this report were collected by the Florida Cancer Data System (Miami, Florida) under contract with the Florida Department of Health, Tallahassee, Florida. The views expressed herein are solely those of the authors and do not necessarily reflect those of the FCDC or FDOH. Cancer incidence data from Louisiana were collected by the Louisiana Tumor Registry, Louisiana State University Health Sciences Center School of Public Health, New Orleans, Louisiana. Cancer incidence data from New Jersey were collected by the New Jersey State Cancer Registry, The Rutgers Cancer Institute of New Jersey, New Brunswick, New Jersey. Cancer incidence data from North Carolina were collected by the North Carolina Central Cancer Registry, Raleigh, North Carolina. Cancer incidence data from Pennsylvania were supplied by the Division of Health Statistics and Research, Pennsylvania Department of Health, Harrisburg, Pennsylvania. The Pennsylvania Department of Health specifically disclaims responsibility for any analyses, interpretations, or conclusions. Cancer incidence data from Arizona were collected by the Arizona Cancer Registry, Division of Public Health Services, Arizona Department of Health Services, Phoenix, Arizona. Cancer incidence data from Texas were collected by the Texas Cancer Registry, Cancer Epidemiology and Surveillance Branch, Texas Department of State Health Services, Austin, Texas. Cancer incidence data from Nevada were collected by the Nevada Central Cancer Registry, Division of Public and Behavioral Health, State of Nevada Department of Health and Human Services, Carson City, Nevada. We are indebted to the participants in the NIH-AARP Diet and Health Study for their outstanding cooperation. We also thank Sigurd Hermansen and Kerry Grace Morrissey from Westat for study outcomes ascertainment and management and Leslie Carroll at Information Management Services for data support and analysis.

Funding Intramural Research Program of the National Institutes of Health (NIH), National Cancer Institute (NCI).

Compliance with ethical standards

Conflict of interest All authors declare that they have no potential conflicts of interest.

References

1. Bouvard V, Loomis D, Guyton KZ, Grosse Y, Ghissassi FE, Benbrahim-Tallaa L, Guha N, Mattock H, Straif K, International Agency for Research on Cancer Monograph Working Group (2015) Carcinogenicity of consumption of red and processed meat. Lancet Oncol 16:1599–1600
2. Daniel CR, Cross AJ, Graubard BI, Park Y, Ward MH, Rothman N, Hollenbeck AR, Chow WH, Sinha R (2012) Large prospective investigation of meat intake, related mutagens, and risk of renal cell carcinoma. Am J Clin Nutr 95:155–162
3. Ferrucci LM, Sinha R, Ward MH, Graubard BI, Hollenbeck AR, Kilfoy BA, Schatzkin A, Michaud DS, Cross AJ (2010) Meat and components of meat and the risk of bladder cancer in the NIH-AARP diet and health study. Cancer 116:4345–4353
4. Stolzenberg-Solomon RZ, Cross AJ, Silverman DT, Schairer C, Thompson FE, Kipnis V, Subar AF, Hollenbeck A, Schatzkin A, Sinha R (2007) Meat and meat-mutagen intake and pancreatic cancer risk in the NIH-AARP cohort. Cancer Epidemiol Biomarkers Prev 16:2664–2675
5. Cross AJ, Ferrucci LM, Risch A, Graubard BI, Ward MH, Park Y, Hollenbeck AR, Schatzkin A, Sinha R (2010) A large prospective study of meat consumption and colorectal cancer risk: an investigation of potential mechanisms underlying this association. Cancer Res 70:2406–2414
6. Flood A, Velie EM, Sinha R, Chaterjee N, Lacey JV Jr, Schairer C, Schatzkin A (2003) Meat, fat, and their subtypes as risk factors for colorectal cancer in a prospective cohort of women. Am J Epidemiol 158:59–68
7. Etemadi A, Abnet CC, Graubard BI, Beane-Freeman L, Freedman ND, Liao L, Dawsey SM, Sinha R (2018) Anatomical subsite can modify the association between meat and meat compounds and risk of colorectal adenocarcinoma: findings from three large U.S. cohorts. Int J Cancer 143:2261–2270
8. Pedersen AN, Kondrup J, Borsheim E (2013) Health effects of protein intake in healthy adults: a systematic literature review. Food Nutr Res 57:21245
9. Lai R, Bian Z, Lin H, Ren J, Zhou H, Guo H (2017) The association between dietary protein intake and colorectal cancer risk: a meta-analysis. World J Surg Oncol 15:169
10. Malik VS, Li Y, Tobias DK, Pan A, Hu FB (2016) Dietary protein intake and risk of type 2 diabetes in US men and women. Am J Epidemiol 183:715–728
11. Song M, Fung TT, Hu FB, Willett WC, Longo VD, Chan AT, Giovannucci EL (2016) Association of animal and plant protein intake with all-cause and cause-specific mortality. JAMA Intern Med 176:1453–1463
12. Schatzkin A, Subar AF, Thompson FE, Harlan LC, Tangrea J, Hollenbeck AR, Hurwitz PE, Coyle L, Schussler N, Michaud DS, Freedman LS, Brown CC, Midthune D, Kipnis V (2001) Design and serendipity in establishing a large cohort with wide dietary intake distributions: the National Institutes of Health-American Association of Retired Persons Diet and Health Study. Am J Epidemiol 154:1119–1125
13. Thompson FE, Kipnis V, Midthune D, Freedman LS, Carroll RJ, Subar AF, Brown CC, Butcher MS, Mouw T, Leitzmann M, Schatzkin A (2008) Performance of a food-frequency questionnaire in the US NIH-AARP (National Institutes of Health-American Association of Retired Persons) diet and health study. Public Health Nutr 11:183–195
14. Michaud D, Midthune D, Hermansen S, Leitzmann M, Harlan L, Kipnis V, Schatzkin A (2005) Comparison of cancer registry case ascertainment with SEER estimates and self-reporting in a subset of the NIH-aarp diet and health study. J Regist Manag 32:70–75

15. Willett WC, Howe GR, Kushi LH (1997) Adjustment for total energy intake in epidemiologic studies. Am J Clin Nutr 65:S1220–S1228 **(Discussion 9S–31S)**
16. Kulldorff M, Sinha R, Chow WH, Rothman N (2000) Comparing odds ratios for nested subsets of dietary components. Int J Epidemiol 29:1060–1064
17. Pasiakos SM, Agarwal S, Lieberman HR, Fulgoni VL 3rd (2015) Sources and amounts of animal, dairy, and plant protein intake of US adults in 2007–2010. Nutrients. 7:7058–7069
18. Pietinen P, Malila N, Virtanen M, Hartman TJ, Tangrea JA, Albanes D, Virtamo J (1999) Diet and risk of colorectal cancer in a cohort of Finnish men. Cancer Causes Control 10:387–396
19. Preis SR, Stampfer MJ, Spiegelman D, Willett WC, Rimm EB (2010) Dietary protein and risk of ischemic heart disease in middle-aged men. Am J Clin Nutr 92:1265–1272
20. Shang X, Scott D, Hodge AM, English DR, Giles GG, Ebeling PR, Sanders KM (2016) Dietary protein intake and risk of type 2 diabetes: results from the Melbourne Collaborative Cohort Study and a meta-analysis of prospective studies. Am J Clin Nutr 104:1352–1365
21. Iacopetta B (2002) Are there two sides to colorectal cancer? Int J Cancer 101:403–408
22. Bufill JA (1990) Colorectal cancer: evidence for distinct genetic categories based on proximal or distal tumor location. Ann Intern Med 113:779–788
23. Halkjaer J, Olsen A, Overvad K, Jakobsen MU, Boeing H, Buijsse B, Palli D, Tognon G, Du H, Forouhi NG, Wareham NJ, Feskens EJ, Sorensen TI, Tjonneland A (2011) Intake of total, animal and plant protein and subsequent changes in weight or waist circumference in European men and women: the diogenes project. Int J Obes 35:1104–1113
24. Corpet DE, Yin Y, Zhang XM, Remesy C, Stamp D, Medline A, Thompson L, Bruce WR, Archer MC (1995) Colonic protein fermentation and promotion of colon carcinogenesis by thermolyzed casein. Nutr Cancer 23:271–281
25. Windey K, De Preter V, Verbeke K (2012) Relevance of protein fermentation to gut health. Mol Nutr Food Res 56:184–196
26. Kim E, Coelho D, Blachier F (2013) Review of the association between meat consumption and risk of colorectal cancer. Nutr Res 33:983–994
27. U.S. Department of Agriculture, Agricultural Research Service (2017) Food patterns equivalents intakes from food: mean amounts consumed per individual, by gender and age, what we eat in America, NHANES 2013–2014. www.ars.usda.gov/nea/bhnrc/fsrg