Received: 3 September 2021 | Revised: 13 October 2021 | Accepted: 19 October 2021

DOI: 10.1111/cas.15196

ORIGINAL ARTICLE

Cancer Science WILEY

Association of *Escherichia coli* containing polyketide synthase in the gut microbiota with colorectal neoplasia in Japan

Motoki Iwasaki[1] | Rieko Kanehara[1] | Taiki Yamaji[1] | Ryoko Katagiri[1] | Michihiro Mutoh[2] | Yuta Tsunematsu[3] | Michio Sato[3] | Kenji Watanabe[3] | Koji Hosomi[4] | Yasuo Kakugawa[5,6] | Hiroaki Ikematsu[7] | Kinichi Hotta[8] | Jun Kunisawa[4] | Keiji Wakabayashi[9] | Takahisa Matsuda[5,6]

[1]Division of Epidemiology, Center for Public Health Sciences, National Cancer Center, Tokyo, Japan

[2]Division of Prevention, Center for Public Health Sciences, National Cancer Center, Tokyo, Japan

[3]Department of Pharmaceutical Sciences, University of Shizuoka, Shizuoka, Japan

[4]Laboratory of Vaccine Materials, Center for Vaccine and Adjuvant Research, and Laboratory of Gut Environmental System, National Institutes of Biomedical Innovation, Health and Nutrition (NIBIOHN), Ibaraki, Japan

[5]Endoscopy Division, National Cancer Center Hospital, Tokyo, Japan

[6]Cancer Screening Center, National Cancer Center Hospital, Tokyo, Japan

[7]Department of Gastroenterology and Endoscopy, National Cancer Center Hospital East, Kashiwa, Japan

[8]Division of Endoscopy, Shizuoka Cancer Center, Shizuoka, Japan

[9]School of Food and Nutritional Sciences, University of Shizuoka, Shizuoka, Japan

Correspondence
Motoki Iwasaki, Division of Epidemiology, Center for Public Health Sciences, National Cancer Center, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan.
Email: moiwasak@ncc.go.jp

Fundinginformation
National Cancer Center Research and Development Fund, Grant/Award Number: 25-A-14, 27-A-5, 30-A-16; Japan Agency for Medical Research and Development (AMED), Grant/Award Number: 16ck0106243h0001, 19ck0106475h0001, 20ck0106551h0001, 20gm1010006h0004; Ministry of Health and Welfare of Japan and Public/Private R&D Investment Strategic Expansion PrograM: PRISM, Grant/Award Number: 20AC5004; SECOM Science and Technology Foundation; Kobayashi Foundation for Cancer Research

Abstract

Escherichia coli containing polyketide synthase in the gut microbiota (*pks*$^+$ *E coli*) produce a polyketide-peptide genotoxin, colibactin, and are suspected to play a role in the development of colorectal neoplasia. To clarify the role of *pks*$^+$ *E coli* in the early stage of tumorigenesis, we investigated whether the *pks* status of *E coli* was associated with the prevalence of colorectal neoplasia. This cross-sectional analysis of data from a prospective cohort in Izu Oshima, Japan included asymptomatic residents aged 40-79 years who underwent screening colonoscopy and provided a stool sample. We identified 543 participants with colorectal neoplasia (22 colorectal cancer and 521 adenoma) as cases and 425 participants with normal colon as controls. The *pks* status of *E coli* was assayed using stool DNA and specific primers that detected *pks*$^+$ *E coli*. The proportion of *pks*$^+$ *E coli* was 32.6% among cases and 30.8% among controls. Compared with those with *pks*$^-$ *E coli*, the odds ratio (OR) (95% confidence interval) for participants with *pks*$^+$ *E coli* was 1.04 (0.77-1.41) after adjusting for potential confounders. No statistically significant associations were observed regardless of tumor site or number of colorectal adenoma lesions. However, stratified analyses revealed increased ORs among participants who consumed cereals over the median intake or vegetables under the median intake. Overall, we found no statistically significant

association between *pks*$^+$ *E coli* and the prevalence of colorectal adenoma lesions among this Japanese cohort. However, positive associations were suggested under certain intake levels of cereals or vegetables.

KEYWORDS

colibactin, colorectal neoplasia, epidemiology, *Escherichia coli*, *pks* island

1 | INTRODUCTION

Colorectal cancer is the third most common cancer worldwide.[1] The majority of cases are sporadic and arise through the traditional adenoma-carcinoma pathway.[2] Accumulating epidemiological evidence indicates the important role of lifestyle and environmental factors in the development of colorectal neoplasia but its etiology is not fully understood.[2,3]

Recently, attention has focused on a potential role of the gut microbiota in colorectal carcinogenesis.[4,5] *Escherichia coli* from the B2 phylogenetic group possesses a genomic island named polyketide synthetase (*pks*), which is thought to produce a polyketide-peptide genotoxin, colibactin. *E coli* containing *pks* (*pks*$^+$ *E coli*) has been shown to induce DNA double-strand breaks, cell cycle arrest, mutations, and chromosomal instability in eukaryotic cells.[6-9] Colibactin also alkylates DNA in vivo and DNA adducts have been identified in mammalian cells and mice exposed to *pks*$^+$ *E coli*.[10] Five studies have compared the prevalence of *pks*$^+$ *E coli* between patients with and without colorectal neoplasia but findings are inconsistent:[11-15] three found significantly higher prevalence among colorectal cancer cases than the control group,[11-13] whereas two showed no statistically significant difference.[14,15] Two of these studies examined the prevalence of colorectal adenoma and observed no statistically significant difference, although one showed a higher prevalence of colorectal adenoma cases.[13,14] However, these studies included a relatively small number of colorectal neoplasia cases and did not adjust for potential confounders.

Here, to better understand the role of *pks*$^+$ *E coli* in the early stage of the adenoma-carcinoma sequence, we investigated whether the *pks* status of *E coli* was associated with the prevalence of colorectal neoplasia. The study was carried out under a cross-sectional design using data from a prospective cohort in Izu Oshima, Japan, which included 543 cases (22 colorectal cancer and 521 adenoma cases) and 425 controls. We also tested the hypothesis that lifestyle and dietary factors modify the association between *pks*$^+$ *E coli* and the prevalence of colorectal neoplasia.

2 | MATERIALS AND METHODS

2.1 | Study cohort

The Oshima study was carried out under a prospective cohort design in Izu Oshima, a small island near the mainland Japanese island of Honshu. The study aimed to evaluate the diagnostic ability and effectiveness of colorectal cancer screening techniques and biomarkers.[16] We recruited all island residents aged 40-79 years without uncontrollable complications, including unstable angina, acute myocardial infarction, heart failure, chronic respiratory disease, and bleeding tendency, which would hinder the safe performance of colonoscopy. The baseline survey, including a self-administered questionnaire survey, blood and stool sample collection, 2-day fecal immunochemical test, and screening colonoscopy, was undertaken between November 2015 and June 2017. Of 4645 residents, 1367 provided written informed consent. This study was approved by the institutional review board of the National Cancer Center, Tokyo, Japan.

2.2 | Questionnaire survey

All participants were asked to complete a self-administered questionnaire before the screening colonoscopy. The questionnaire enquired about lifestyle factors, such as personal medical history, present medication, family history of cancer, cigarette smoking, alcohol drinking, and physical activity, among others. It also included a food frequency questionnaire (FFQ). The FFQ was originally used in the Japan Public Health Center-based Prospective Study for the Next Generation (JPHC-NEXT Study), and contained an added item, *kusaya*, a dried fish that is popular in Izu Oshima. The original FFQ was validated in middle-aged and elderly Japanese using 12-day weighed food records (3 days per season).[17] It consists of 67 food and beverage items with nine frequency categories and standard portions/units, and asks about the usual consumption of listed foods during the previous year. Frequency response choices for food items are less than once per month, 1-3 times per month, 1-2 times per week, 3-4 times per week, 5-6 times per week, once per day, 2-3 times per day, 4-6 times per day, and 7 or more times per day. Standard portion sizes are specified for each food item in the three "amount" choices of small (50% smaller than standard), medium (standard), and large (50% larger). Daily food intake is calculated by multiplying frequency by standard portion and relative size for each food item. Intake of energy and nutrients is calculated using the Standard Tables of Food Composition in Japan 2015.[18]

2.3 | Stool sample collection and laboratory analysis

Stool sample collection vials containing 3 ml GuSCN solution (TechnoSuruga Laboratory Co., Ltd) along with information about

the collection procedure were sent to participants. The sample was collected by the participant prior to preparation for the colonoscopy procedure and stored at room temperature until the colonoscopy procedure. The vials were then stored at −80°C until analysis. Stool DNA was extracted from a portion of frozen stool by the bead beating method, as detailed elsewhere.[19]

To confirm that the *E coli* was a *pks*$^+$ strain, PCR was carried out to amplify genes from the *clb* cluster using bacterial genomic DNA as a template. The details have been reported elsewhere.[20,21] In brief, two primer sets were used to amplify each of the genes in the cluster, namely clbB-F/clbB-R for *clbB* and clbQ-F/clbQ-R for *clbQ*. Participants for whom *clbB* and *clbQ* were unambiguously detected from feces were defined as *pks*$^+$ *E coli* individuals.

2.4 | Colonoscopy procedure

All colonoscopy procedures were undertaken to examine the whole colon and rectum using video colonoscopes with a magnification function (CF-HQ290ZI, PCF-Q260AZI; Olympus Co.). A total of 25 experienced endoscopists who were board-certified by the Japanese Gastrointestinal Endoscopy Society participated in the study and carried out the colonoscopies. Polyethylene glycol or magnesium citrate solution was given in the morning of the day of the procedure for bowel preparation.

2.5 | Selection of cases and control

Of the 1367 participants, we excluded participants who did not undergo colonoscopy, underwent incomplete examination (cecum not reached in colonoscopy), had a history of cancer, colorectal polyp, colorectal surgery, or colonoscopic treatment based on a self-administered questionnaire, or who did not provide a stool sample. We further excluded participants who reported extreme energy intakes (below the 2.5 or over the 97.5 percentiles), leaving 1034 participants. Among these, 22 participants had colorectal cancer, 521 had one or more adenomas, 49 had hyperplastic polyp only, and 17 had other lesions (eg, neuroendocrine tumor, nonneoplastic lesion) based on a pathologically confirmed diagnosis. The remaining 425 had a normal colon. After exclusion of participants with hyperplastic polyp only or other lesions, we considered the 543 participants with colorectal neoplasia (colorectal cancer or adenoma) as cases and the 425 participants with normal colon as controls. Additionally, we defined advanced colorectal neoplasia as comprising colorectal cancer and advanced adenoma (adenoma with a diameter of 10 mm or more, high-grade dysplasia, or prominent villous component).[22,23] Subsites of colon neoplasia were defined by a location in the proximal colon (cecum and ascending and transverse colon) or distal colon (descending and sigmoid colon).

2.6 | Statistical analysis

Dietary intakes of food groups and nutrients were energy-adjusted by the residual regression method. Case-control comparisons for mean, median, and proportions were tested with the *t* test, Wilcoxon rank-sum test, and χ^2 test, respectively. An unconditional logistic regression model was used to estimate odds ratio (OR) and 95% confidence intervals (CI) of the prevalence of colorectal neoplasia according to the *pks* status of *E coli*. The regression models were adjusted for age (continuous), sex, cigarette smoking (never smokers, past smokers, and <20, 20-39, ≥40 pack-years for current smokers), alcohol consumption (nondrinkers, past drinkers, occasional drinkers, and <150, 150-299, 300-449, ≥450 g/wk for regular drinkers), body mass index (kg/m^2) (<21, 21-23.9, 24-26.9, 27-29.9, ≥30), physical activity (metabolic equivalent-h/d, quartile category), family history of colorectal cancer, nonsteroidal anti-inflammatory drug use, and energy-adjusted intakes of cereals, vegetables, fruits, meats, and dairy products (quartile category). Stratified analyses were undertaken according to dichotomous categories of cigarette smoking, alcohol consumption, body mass index, physical activity, and energy-adjusted intakes of cereals, vegetables, fruits, meats, and dairy products. An interaction term was created by multiplying variables for *pks* status by those for dichotomous categories of each stratified variable, and its significance was statistically evaluated by the likelihood ratio test with 1 *df*.

In order to clarify factors associated with the prevalence of *pks*$^+$ *E coli*, risk factors of colorectal cancer and dietary intakes of food groups and nutrients were compared between participants with and without *pks*$^+$ *E coli* among the control group. Comparisons in mean, median, and proportions were tested with the *t* test, Wilcoxon rank-sum test, and χ^2 test, respectively. Furthermore, an unconditional logistic regression model was used to estimate ORs and 95% CIs of *pks*$^+$ *E coli* participants according to risk factors of colorectal cancer and quartile categories of energy-adjusted dietary intake. Linear trends for ORs in the logistic regression model were tested using the quartile categories as ordinal variables.

All reported *p* values are two-sided, and significance level was set at $P < .05$. All statistical analyses were undertaken using SAS 9.4 (SAS Institute Inc.).

3 | RESULTS

Table 1 presents participant characteristics by case-control status. The proportion of men was higher in cases than controls and cases were older, smoked more, and consumed more alcoholic beverages than controls. However, cases consumed fewer dairy products than controls. The distribution of other variables including body mass index, physical activity, and dietary intake except dairy products was similar between cases and controls.

We examined factors associated with the prevalence of *pks*$^+$ *E coli* among the control group (Tables S1–S6). No statistically

TABLE 1 Characteristics of study participants with colorectal neoplasia (cases) or normal colon (controls)

	Case		Control		*P* value
Number	543		425		
Men, n (%)	274	(50.5)	147	(34.6)	<.010
Age, years; mean (SD)	62.8	(9.8)	58.1	(11.3)	<.010
Body mass index, kg/m^2; mean (SD)	23.2	(3.3)	23.2	(3.5)	.850
Physical activity, metabolic equivalent-hours/day; mean (SD)	40.8	(6.9)	40.5	(5.8)	.530
Current smokers, n (%)	115	(21.6)	58	(13.8)	<.010
Alcohol intake, ≥1 d/wk; n (%)	277	(51.1)	170	(40.2)	<.010
Family history of colorectal cancer, yes; n (%)	60	(11.1)	39	(9.2)	.340
Nonsteroidal anti-inflammatory drug use, yes; n (%)	51	(9.4)	48	(11.3)	.330
Dietary intake, median (interquartile range)					
Energy, kcal/d	1418.8	(1076.2-1792.7)	1328.0	(1050.5-1752.0)	.230
Cereals, g/d	389.3	(296.4-507.1)	376.9	(296.7-483.6)	.400
Vegetables, g/d	81.9	(47.3-136.0)	90.0	(49.3-139.3)	.240
Fruits, g/d	48.7	(13.2-105.5)	52.0	(13.7-119.8)	.280
Meats, g/d	36.4	(17.9-64.8)	35.3	(19.4-67.0)	.770
Dairy products, g/d	42.9	(0-157.1)	50.0	(0-200.0)	.045

significant association was found for selected risk factors of colorectal cancer (Tables S1 and S2). Comparison of median intakes of food groups and nutrients showed no statistically significant difference between participants with and without *pks*$^+$ *E coli* among the control group except with regard to cruciferous vegetable intake (Tables S3 and S4): control participants with *pks*$^+$ *E coli* had significantly higher intake of cruciferous vegetables than those without, and energy-adjusted cruciferous vegetable intake was significantly associated with a higher prevalence of *pks*$^+$ *E coli* after adjustment for age, sex, and other lifestyle factors (Tables S3 and S5). In addition, energy-adjusted vitamin C intake was significantly associated with a higher prevalence of *pks*$^+$ *E coli*, whereas energy-adjusted chromium intake was significantly associated with a lower prevalence (Table S6).

The proportion of *pks*$^+$ *E coli* was 32.6% among cases and 30.8% among controls. Table 2 shows ORs of the prevalence of colorectal neoplasia according to *pks* status. Compared with the participants with *pks*$^-$ *E coli*, the OR (95% CI) for participants with *pks*$^+$ *E coli* was 1.04 (0.77-1.41) after adjusting for potential confounding factors. Although inclusion of dietary factors as potential confounding factors did not change the result, the same result was also obtained by further sensitivity analysis in which meat intake was replaced with red meat intake (data not shown). No significant association was observed regardless of sex and age group. Furthermore, site-specific analysis in 238 proximal, 179 distal, and 69 rectal cases following exclusion of 57 unclassified cases due to multiple lesions found no significant association regardless of site.

We reclassified colorectal neoplasia cases into colorectal cancer (n = 22) and adenoma cases (n = 521) and further defined advanced colorectal neoplasia (n = 102) (Table 3). The proportion of *pks*$^+$ *E coli* was 40.9% for colorectal cancer, 32.2% for adenoma, and 26.5% for advanced colorectal neoplasia cases. Although no significant association was found for each outcome, the OR (95% CI) of colorectal cancer was 1.44 (0.48-4.26) among participants with *pks*$^+$ *E coli*. We further divided adenoma cases according to the number of adenoma lesions (Table 4). The *pks* status of *E coli* was not significantly associated with the prevalence of colorectal adenoma regardless of the number of lesions.

Stratified analyses by selected risk factors for colorectal cancer are shown in Table 5. No significant interaction was observed for cigarette smoking, alcohol consumption, body mass index, physical activity, or energy-adjusted intakes of fruits, meats, and dairy products. However, *pks*$^+$ *E coli* was significantly associated with a higher prevalence of colorectal neoplasia among participants who consumed cereals over the median intake. A nonsignificant inverse association was observed among participants who consumed cereals under the median intake. A statistically significant interaction was found between *pks* status and energy-adjusted cereal intake on the prevalence of colorectal neoplasia ($P = .002$). In addition, a marginally nonsignificant interaction was found for stratified analysis by energy-adjusted vegetable intake ($P = .08$). A positive association was observed among participants who consumed vegetables under the median intake, whereas an inverse association was seen among those who consumed over the median intake. Further stratified analyses by rice, pickled vegetables, green and yellow vegetables, cruciferous vegetables, red meats, and total dietary fiber intake revealed a statistically significant interaction for cruciferous vegetable intake ($P = .01$) (Table S7).

TABLE 2 Association of *Escherichia coli* containing polyketide synthase (*pks*$^+$ *E coli*) with the prevalence of colorectal neoplasia (colorectal cancer or adenoma)

	pks$^-$ *E coli*	*pks*$^+$ *E coli*
All subjects		
No. of cases	366	177
No. of controls	294	131
OR (95% CI)[a]	1 (ref.)	1.04 (0.78-1.39)
OR (95% CI)[b]	1 (ref.)	1.04 (0.77-1.41)
OR (95% CI)[c]	1 (ref.)	1.04 (0.77-1.41)
Sex		
Men		
No. of cases	183	91
No. of controls	104	43
OR (95% CI)[c]	1 (ref.)	1.01 (0.61-1.67)
Women		
No. of cases	183	86
No. of controls	190	88
OR (95% CI)[c]	1 (ref.)	1.01 (0.68-1.50)
Age group		
40-49 y		
No. of cases	47	20
No. of controls	85	38
OR (95% CI)[c]	1 (ref.)	0.63 (0.29-1.40)
50-59 y		
No. of cases	77	33
No. of controls	69	32
OR (95% CI)[c]	1 (ref.)	1.33 (0.64-2.73)
60-69 y		
No. of cases	142	72
No. of controls	90	32
OR (95% CI)[c]	1 (ref.)	1.42 (0.80-2.50)
70-79 y		
No. of cases	100	52
No. of controls	50	29
OR (95% CI)[c]	1 (ref.)	0.67 (0.33-1.37)
Tumor location		
Proximal colon		
No. of cases	156	82
No. of controls	294	131
OR (95% CI)[c]	1 (ref.)	1.14 (0.78-1.67)
Distal colon		
No. of cases	125	54
No. of controls	294	131
OR (95% CI)[c]	1 (ref.)	1.06 (0.69-1.62)
Rectum		
No. of cases	42	27
No. of controls	294	131
OR (95% CI)[c]	1 (ref.)	1.35 (0.75-2.43)

(Continues)

Abbreviations: CI, confidence interval; OR, odds ratio; ref., reference.

[a]Adjusted for sex and age (continuous).

[b]Further adjusted for cigarette smoking (never smokers, past smokers, <20, 20-39, ≥40 pack-years for current smokers), alcohol consumption (nondrinkers, past drinkers, occasional drinkers, <150, 150-299, 300-449, ≥450 g ethanol/wk for regular drinkers), body mass index (kg/m^2) (<21, 21-23.9, 24-26.9, 27-29.9, ≥30), physical activity (metabolic equivalent-h/d, quartile category), family history of colorectal cancer (yes, no), and nonsteroidal anti-inflammatory drug use (yes, no).

[c]Further adjusted for energy-adjusted intakes of cereals, vegetables, fruits, meats, and dairy products (quartile category).

4 | DISCUSSION

Overall, we found no statistically significant association between *pks*$^+$ *E coli* and the prevalence of colorectal adenoma lesions in the present cross-sectional analysis. This finding suggests that *pks*$^+$ *E coli* might not play a role in the early stage of the adenoma-carcinoma sequence. Stratified analyses by selected risk factors for colorectal cancer revealed effect modification by cereal and vegetable intake: a positive association was suggested only among participants who consumed cereals over the median intake or vegetables under the median intake.

Our findings, based on 521 colorectal adenoma cases, are consistent with a Japanese study that showed no statistically significant difference in the prevalence of *pks*$^+$ *E coli* between 37 colorectal adenoma cases (51%) and 26 controls (46%).[14] They are also consistent with a Swedish study that found a higher prevalence of *pks*$^+$ *E coli* among 134 colorectal adenoma cases (31.3%) than 65 controls (18.5%), although the difference was not statistically significant.[13] A statistically significant high prevalence of *pks*$^+$ *E coli* was observed among 25 patients with familial adenomatous polyposis (68%) compared to 23 controls (22%),[24] although this might rather suggest a different role of *pks*$^+$ *E coli* in the development of sporadic versus hereditary colorectal cancer.

Regarding the prevalence of *pks*$^+$ *E coli* among colorectal cancer cases, three studies from the UK, France, and Sweden found a statistically significant high prevalence compared to controls.[11-13] These studies involved 21, 38, and 39 cases, respectively, and had a prevalence between cases and controls of 67% and 21%, 55% and 19%, and 56.4% and 18.5%, respectively.[11-13] In contrast, no statistically significant difference in the prevalence of *pks*$^+$ *E coli* was observed between 35 colorectal cancer cases (43%) and 26 controls (46%) in Japan.[14] In addition, a study from Malaysia showed higher prevalence among 48 colorectal cancer cases (16.7%) than 23 controls (4.3%) but without statistical significance.[15] The two studies from Asian countries are in general agreement with our findings; however, our study included only 22 patients with colorectal cancer. Although the reason for these discrepant findings is unclear, one possible

TABLE 3 Association of *Escherichia coli* containing polyketide synthase (*pks*$^+$ *E coli*) with the prevalence of colorectal cancer, adenoma, and advanced colorectal neoplasia[a]

	pks$^-$ *E coli*	*pks*$^+$ *E coli*
Colorectal cancer		
No. of cases	13	9
No. of controls	294	131
OR (95% CI)[b]	1 (ref.)	1.44 (0.48-4.26)
Adenoma		
No. of cases	353	168
No. of controls	294	131
OR (95% CI)[b]	1 (ref.)	1.04 (0.76-1.41)
Advanced colorectal neoplasia[a]		
No. of cases	75	27
No. of controls	294	131
OR (95% CI)[b]	1 (ref.)	0.85 (0.48-1.50)

Abbreviations: CI, confidence interval; OR, odds ratio; ref., reference.

[a]Advanced colorectal neoplasia comprised colorectal cancer and advanced adenoma (adenoma with a diameter ≥10 mm, high-grade dysplasia, or prominent villous component).

[b]Adjusted for sex, age (continuous), cigarette smoking (never smokers, past smokers, <20, 20-39, ≥40 pack-years for current smokers), alcohol consumption (nondrinkers, past drinkers, occasional drinkers, <150, 150-299, 300-449, ≥450 g ethanol/wk for regular drinkers), body mass index (kg/m^2) (<21, 21-23.9, 24-26.9, 27-29.9, ≥30), physical activity (metabolic equivalent-h/d, quartile category), family history of colorectal cancer (yes, no), nonsteroidal anti-inflammatory drug use (yes, no), and energy-adjusted intakes of cereals, vegetables, fruits, meats, and dairy products (quartile category).

TABLE 4 Association of *Escherichia coli* containing polyketide synthase (*pks*$^+$ *E coli*) with the prevalence of colorectal adenoma according to number of lesions

	pks$^-$ *E coli*	*pks*$^+$ *E coli*
One adenoma		
No. of cases	162	90
No. of controls	294	131
OR (95% CI)[a]	1 (ref.)	1.21 (0.85-1.73)
Two adenomas		
No. of cases	86	39
No. of controls	294	131
OR (95% CI)[a]	1 (ref.)	1.15 (0.69-1.92)
Three or four adenomas		
No. of cases	69	25
No. of controls	294	131
OR (95% CI)[a]	1 (ref.)	0.71 (0.38-1.31)
More than five adenomas		
No. of cases	36	14
No. of controls	294	131
OR (95% CI)[a]	1 (ref.)	0.97 (0.42-2.24)

Abbreviations: CI, confidence interval; OR, odds ratio; ref., reference.

[a]Adjusted for sex, age (continuous), cigarette smoking (never smokers, past smokers, <20, 20-39, ≥40 pack-years for current smokers), alcohol consumption (nondrinkers, past drinkers, occasional drinkers, <150, 150-299, 300-449, ≥450 g ethanol/wk for regular drinkers), body mass index (kg/m^2) (<21, 21-23.9, 24-26.9, 27-29.9, ≥30), physical activity (metabolic equivalent-h/d, quartile category), family history of colorectal cancer (yes, no), nonsteroidal anti-inflammatory drug use (yes, no), and energy-adjusted intakes of cereals, vegetables, fruits, meats, and dairy products (quartile category).

explanation is that the different prevalence of *pks*$^+$ *E coli* among controls might reflect a difference in microbiota composition across populations and thus a difference in the production of colibactin. Of interest, prevalence in a previous Japanese study (46%) and our present study (30.8%) was higher than in studies from European countries (approximately 20%). If *pks*$^+$ *E coli* among controls in the Japanese studies did not produce sufficient levels of colibactin for any reason, the presence of *pks*$^+$ *E coli* would not necessarily imply colibactin exposure. Accordingly, further studies should examine the association between exposure to colibactin by *pks*$^+$ *E coli* and the prevalence of colorectal neoplasia.

As *pks*$^+$ *E coli* was detected in the gut of newborns, mother to offspring transmission during birth is suspected.[25,26] Nevertheless, factors associated with long-term persistence have not been clarified. Further investigation of factors associated with the prevalence of *pks*$^+$ *E coli* is needed, and might help our understanding of the difference in prevalence of *pks*$^+$ *E coli* among populations. In this study, we found no statistically significant association of age, sex, and selected risk factors of colorectal cancer with the prevalence of *pks*$^+$ *E coli*, suggesting that these are not major factors in the different prevalence of *pks*$^+$ *E coli* among populations, as discussed above. However, we found that energy-adjusted intakes of cruciferous vegetables and vitamin C were significantly associated with higher prevalence, whereas energy-adjusted intake of chromium was significantly associated with lower prevalence. To our knowledge, our study is the second to examine the association between dietary factors and prevalence of *pks*$^+$ *E coli*. The first study, undertaken in middle-aged Japanese, showed an inverse association between intake of green tea and manganese and the prevalence of *pks*$^+$ *E coli*.[27] These dietary factors are not known as established risk factors for colorectal cancer[3,28] but could influence the *pks* status of *E coli*; however, the reported findings are inconsistent. Further accumulation of evidence is needed to clarify factors associated with the prevalence of *pks*$^+$ *E coli*.

Given that some dietary risk factors for colorectal cancer are metabolized by the gut microbiota and might influence its composition,[28] we hypothesized that they could modify the association between the *pks* status of *E coli* and the prevalence of colorectal neoplasia. Indeed, our stratified analyses by cereal, vegetable, and cruciferous vegetable intake found interactions. A stratified analysis by dietary fiber found no statistically significant interaction, although cereals and vegetables are rich in dietary fiber. A positive association was observed among participants who consumed cereals over the median intake and an inverse association was seen among those with

TABLE 5 Association of *Escherichia coli* containing polyketide synthase (*pks*$^+$ *E coli*) with the prevalence of colorectal neoplasia (colorectal cancer or adenoma) according to risk factors of colorectal cancer

	pks$^-$ *E coli*	*pks*$^+$ *E coli*	*P* for interaction
Smoking status			.470
Never smokers			
No. of cases	142	80	
No. of controls	153	76	
OR (95% CI)[a]	1 (ref.)	0.96 (0.63-1.47)	
Ever smokers			
No. of cases	216	94	
No. of controls	137	54	
OR (95% CI)[a]	1 (ref.)	1.17 (0.74-1.84)	
Alcohol intake			.380
Nondrinkers			
No. of cases	153	80	
No. of controls	158	70	
OR (95% CI)[a]	1 (ref.)	1.24 (0.80-1.90)	
Drinkers			
No. of cases	213	96	
No. of controls	135	60	
OR (95% CI)[a]	1 (ref.)	0.87 (0.55-1.36)	
Body mass index (kg/m^2)			.610
<25			
No. of cases	267	127	
No. of controls	215	98	
OR (95% CI)[a]	1 (ref.)	1.14 (0.80-1.63)	
≥25			
No. of cases	93	44	
No. of controls	74	31	
OR (95% CI)[a]	1 (ref.)	0.96 (0.49-1.89)	
Physical activity (metabolic equivalent-h/d)			.310
<Median			
No. of cases	176	104	
No. of controls	142	69	
OR (95% CI)[a]	1 (ref.)	1.22 (0.79-1.88)	
≥Median			
No. of cases	190	73	
No. of controls	150	62	
OR (95% CI)[a]	1 (ref.)	0.88 (0.55-1.39)	
Energy-adjusted cereal intake[b]			.002
<Median			
No. of cases	201	75	
No. of controls	140	72	
OR (95% CI)[a]	1 (ref.)	0.66 (0.43-1.03)	
≥Median			
No. of cases	165	102	
No. of controls	154	59	

(Continues)

TABLE 5 (Continued)

	pks⁻ *E coli*	*pks*⁺ *E coli*	*P* for interaction
OR (95% CI)[a]	1 (ref.)	1.66 (1.05-2.61)	
Energy-adjusted vegetable intake[b]			.080
<Median			
No. of cases	195	97	
No. of controls	155	57	
OR (95% CI)[a]	1 (ref.)	1.33 (0.84-2.10)	
≥Median			
No. of cases	171	80	
No. of controls	139	74	
OR (95% CI)[a]	1 (ref.)	0.81 (0.52-1.26)	
Energy-adjusted fruit intake[b]			.210
<Median			
No. of cases	203	82	
No. of controls	149	63	
OR (95% CI)[a]	1 (ref.)	0.83 (0.53-1.31)	
≥Median			
No. of cases	163	95	
No. of controls	145	68	
OR (95% CI)[a]	1 (ref.)	1.32 (0.84-2.06)	
Energy-adjusted meat intake[b]			.360
<Median			
No. of cases	189	98	
No. of controls	147	65	
OR (95% CI)[a]	1 (ref.)	1.23 (0.80-1.89)	
≥Median			
No. of cases	177	79	
No. of controls	147	66	
OR (95% CI)[a]	1 (ref.)	0.94 (0.59-1.50)	
Energy-adjusted dairy product intake[b]			.720
<Median			
No. of cases	191	104	
No. of controls	141	71	
OR (95% CI)[a]	1 (ref.)	1.10 (0.72-1.70)	
≥Median			
No. of cases	175	73	
No. of controls	153	60	
OR (95% CI)[a]	1 (ref.)	1.00 (0.63-1.58)	

Abbreviations: CI, confidence interval; OR, odds ratio; ref., reference.

[a]Adjusted for sex, age (continuous), cigarette smoking (never smokers, past smokers, <20, 20-39, ≥40 pack-years for current smokers), alcohol consumption (nondrinkers, past drinkers, occasional drinkers, <150, 150-299, 300-449, ≥450 g ethanol/wk for regular drinkers), body mass index (kg/m^2) (<21, 21-23.9, 24-26.9, 27-29.9, ≥30), physical activity (metabolic equivalent-h/d, quartile category), family history of colorectal cancer (yes, no), nonsteroidal anti-inflammatory drug use (yes, no), and energy-adjusted intakes of cereals, vegetables, fruits, meats, and dairy products (quartile category).

[b]Cereal intake includes seven items: cooked rice, grain, millet, sawa millet, bread [including pastry], udon, soba, and brown rice. Vegetable intake includes 15 items: carrot, spinach, pumpkin, salted pickles of Chinese radish, green leafy vegetables, Chinese cabbage, cucumber, and eggplant, tomato, Welsh onion, garland chrysanthemum, broccoli, onion, Chinese cabbage, and tomato juice. Fruit intake includes six items: mandarin orange, apple, watermelon, banana, orange juice, and salted pickles of plum. Meat intake includes 12 items: steaks, grilled beef, stir-fried pork, stewed beef, stir-fried pork, deep-fried pork, stewed pork, western-style stewed pork, Japanese-style pork in soup, pork liver, deep-fried chicken, and chicken liver. Dairy product intake includes two items: whole milk and low-fat milk.

consumption under the median intake. The opposite direction of associations was observed for vegetable and cruciferous vegetable intake. The reasons for the positive association among participants who consumed cereals over the median intake or vegetables under the median intake, particularly cruciferous vegetables, are unclear. As dietary fiber did not modify the association in this study, factors other than dietary fiber might be considered. Cereals are also high glycemic index foods and major sources of dietary carbohydrates, including starch, which have been associated with chronic conditions such as obesity and diabetes.[29,30] Thus, we speculate that a high cereal or low vegetable intake might imply a somewhat unhealthy diet, and an intestinal environment characterized by relatively high cereal or low vegetable intake might promote carcinogenesis due to *pks*$^+$ *E coli*. As this is the first report of such interaction, confirmation in other studies is warranted.

The strengths of our study include its relatively large number of adenoma cases (n = 521) and adjustment for potential confounders, including dietary factors. In addition, all participants underwent total colonoscopy, reducing the possibility of misclassification of case and control status. Nevertheless, several limitations need to be addressed. First is the cross-sectional design, and the possibility that the observed associations were subject to reverse causality. Second, our study cohort was mainly derived from a single small island in Japan, which might limit our representativeness. Considering that large variation in microbiota composition among individuals is likely, primarily due to different external environmental factors, the generalizability of our finding to other populations is also limited. Finally, although we adjusted for known or potential confounding factors in the multivariable models, residual or unmeasured confounding remains possible.

In conclusion, we found that the *pks* status of *E coli* was not significantly associated with the prevalence of colorectal adenoma lesions in a cross-sectional analysis in a Japanese cohort. However, positive associations were suggested under certain intake level of cereals and vegetables, and thus dietary intake might modify this association. Further studies using an appropriate biomarker of *pks*$^+$ *E coli* exposure from a large number of colorectal neoplasia cases are required.

ACKNOWLEDGMENTS

We are grateful to Drs. Toshio Uraoka, Nozomu Kobayashi, Kenichiro Imai, Sayo Ito, Yasuhiro Oono, Shinsuke Kiriyama, Seiko Tsujie, Naoko Nakano, Taku Sakamoto, Masayoshi Yamada, Masau Sekiguchi, Hiroyuki Takamaru, Seiichiro Abe, Minori Matsumoto, Hajime Takisawa, Shiko Kuribayashi, Teppei Tagawa, Kenichi Konda, Yusaku Tanaka, Daisuke Hihara, Kouji Yamamoto, and Yutaka Saito for their valuable cooperation in performing the colonoscopies for the study participants. We acknowledge the excellent contribution of Ms Mika Mori to the management of the study. This study was supported by the National Cancer Center Research and Development Fund (25-A-14, 27-A-5 and 30-A-16), Practical Research for Innovative Cancer Control, Japan Cancer Research Project, Japan Agency for Medical Research and Development (AMED) (16ck0106243h0001, 19ck0106475h0001 and 20ck0106551h0001), Advanced Research & Development Programs for Medical Innovation, AMED (20gm1010006h0004), the Ministry of Health and Welfare of Japan and Public/Private R&D Investment Strategic Expansion PrograM: PRISM (20AC5004), SECOM Science and Technology Foundation, and Kobayashi Foundation for Cancer Research.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

ORCID

Motoki Iwasaki https://orcid.org/0000-0003-3319-4131
Michihiro Mutoh https://orcid.org/0000-0003-2054-7655
Kenji Watanabe https://orcid.org/0000-0002-0463-4831

REFERENCES

1. Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2018;68:394-424.
2. Dekker E, Tanis PJ, Vleugels JLA, et al. Colorectal cancer. *Lancet*. 2019;394:1467-1480.
3. Keum N, Giovannucci E. Global burden of colorectal cancer: emerging trends, risk factors and prevention strategies. *Nat Rev Gastroenterol Hepatol*. 2019;16:713-732.
4. Mizutani S, Yamada T, Yachida S. Significance of the gut microbiome in multistep colorectal carcinogenesis. *Cancer Sci*. 2020;111:766-773.
5. Vacante M, Ciuni R, Basile F, et al. Gut Microbiota and Colorectal Cancer Development: A Closer Look to the Adenoma-Carcinoma Sequence. *Biomedicines*. 2020;8:489.
6. Nougayrede JP, Homburg S, Taieb F, et al. Escherichia coli induces DNA double-strand breaks in eukaryotic cells. *Science*. 2006;313:848-851.
7. Cuevas-Ramos G, Petit CR, Marcq I, et al. Escherichia coli induces DNA damage in vivo and triggers genomic instability in mammalian cells. *Proc Natl Acad Sci U S A*. 2010;107:11537-11542.
8. Kawanishi M, Hisatomi Y, Oda Y, et al. In vitro genotoxicity analyses of colibactin-producing E. coli isolated from a Japanese colorectal cancer patient. *J Toxicol Sci*. 2019;44:871-876.
9. Zhou T, Hirayama Y, Tsunematsu Y, et al. Isolation of new colibactin metabolites from wild-type *Escherichia coli* and in situ trapping of a mature colibactin derivative. *J Am Chem Soc*. 2021.
10. Wilson MR, Jiang Y, Villalta PW, et al. The human gut bacterial genotoxin colibactin alkylates DNA. *Science*. 2019;363:eaar7785.
11. Arthur JC, Perez-Chanona E, Muhlbauer M, et al. Intestinal inflammation targets cancer-inducing activity of the microbiota. *Science*. 2012;338:120-123.
12. Buc E, Dubois D, Sauvanet P, et al. High prevalence of mucosa-associated *E. coli* producing Cyclomodulin and Genotoxin in Colon Cancer. *PLoS One*. 2013;8(2):e56964.
13. Eklof V, Lofgren-Burstrom A, Zingmark C, et al. Cancer-associated fecal microbial markers in colorectal cancer detection. *Int J Cancer*. 2017;141:2528-2536.
14. Shimpoh T, Hirata Y, Ihara S, et al. Prevalence of pks-positive Escherichia coli in Japanese patients with or without colorectal cancer. *Gut Pathog*. 2017;9:35.
15. Iyadorai T, Mariappan V, Vellasamy KM, et al. Prevalence and association of pks+ Escherichia coli with colorectal cancer in patients at the University Malaya Medical Centre. *Malaysia. Plos One*. 2020;15:e0228217.
16. Sekiguchi M, Kakugawa Y, Ikematsu H, et al. Risk stratification score improves sensitivity for advanced colorectal neoplasia in

colorectal cancer screening: The Oshima Study Workgroup. *Clin Trans Gastroenterol*. 2021;12:e00319.
17. Yokoyama Y, Takachi R, Ishihara J, et al. Validity of short and long self-administered food frequency questionnaires in ranking dietary intake in middle-aged and elderly Japanese in the Japan public health center-based prospective study for the next generation (JPHC-NEXT) protocol area. *J Epidemiol*. 2016;26:420-432.
18. Office for Resources PD, Science and Technology Policy Bureau, Ministry of Education, Culture, Sports, Science and Technology, Japan. Standard Tables of Food Composition in Japan 2015. https://www.mext.go.jp/en/policy/science_technology/policy/title01/detail01/1374030.htm
19. Hosomi K, Ohno H, Murakami H, et al. Method for preparing DNA from feces in guanidine thiocyanate solution affects 16S rRNA-based profiling of human microbiota diversity. *Sci Rep*. 2017;7:4339.
20. Hirayama Y, Tsunematsu Y, Yoshikawa Y, et al. Activity-Based Probe for Screening of High-Colibactin Producers from Clinical Samples. *Org Lett*. 2019;21:4490-4494.
21. Yoshikawa Y, Tsunematsu Y, Matsuzaki N, et al. Characterization of Colibactin-Producing Escherichia coli Isolated from Japanese Patients with Colorectal Cancer. *Jpn J Infect Dis*. 2020;73:437-442.
22. Lieberman DA, Weiss DG, Bond JH, et al. Use of colonoscopy to screen asymptomatic adults for colorectal cancer. Veterans Affairs Cooperative Study Group 380. *N Engl J Med*. 2000;343:162-168.
23. Winawer SJ, Zauber AG. The advanced adenoma as the primary target of screening. *Gastrointest Endosc Clin N Am*. 2002;12(1–9):1-9.
24. Dejea CM, Fathi P, Craig JM, et al. Patients with familial adenomatous polyposis harbor colonic biofilms containing tumorigenic bacteria. *Science*. 2018;359:592-597.
25. Payros D, Secher T, Boury M, et al. Maternally acquired genotoxic Escherichia coli alters offspring's intestinal homeostasis. *Gut Microbes*. 2014;5:313-325.
26. Tsunematsu Y, Hosomi K, Kunisawa J, et al. Mother-to-infant transmission of the carcinogenic colibactin-producing bacteria. *BMC Microbiol*. 2021;21:235.
27. Watanabe D, Murakami H, Ohno H, et al. Association between dietary intake and the prevalence of tumourigenic bacteria in the gut microbiota of middle-aged Japanese adults. *Sci Rep*. 2020;10:15221.
28. Song M, Garrett WS, Chan AT. Nutrients, foods, and colorectal cancer prevention. *Gastroenterology*. 2015;148:1244-1260.
29. Villegas R, Liu S, Gao YT, et al. Prospective study of dietary carbohydrates, glycemic index, glycemic load, and incidence of type 2 diabetes mellitus in middle-aged Chinese women. *Arch Intern Med*. 2007;167:2310-2316.
30. Sluijs I, van der Schouw YT, van der AD, et al. Carbohydrate quantity and quality and risk of type 2 diabetes in the European Prospective Investigation into Cancer and Nutrition-Netherlands (EPIC-NL) study. *Am J Clin Nutr*. 2010;92:905-911.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Iwasaki M, Kanehara R, Yamaji T, et al. Association of *Escherichia coli* containing polyketide synthase in the gut microbiota with colorectal neoplasia in Japan. *Cancer Sci*. 2022;113:277–286. doi:10.1111/cas.15196