

RESEARCH ARTICLE

Open Access



Stool pattern is associated with not only the prevalence of tumorigenic bacteria isolated from fecal matter but also plasma and fecal fatty acids in healthy Japanese adults

Daiki Watanabe¹, Haruka Murakami¹, Harumi Ohno¹, Kumpei Tanisawa¹, Kana Konishi¹, Kikue Todoroki-Mori², Yuta Tsunematsu³, Michio Sato³, Yuji Ogata³, Noriyuki Miyoshi⁴, Naoto Kubota², Jun Kunisawa⁵, Keiji Wakabayashi⁴, Tetsuya Kubota^{2,6,7}, Kenji Watanabe³ and Motohiko Miyachi^{1,8*} 

Abstract

Background: Colibactin-producing *Escherichia coli* containing polyketide synthase (*pks*⁺ *E. coli*) has been shown to be involved in colorectal cancer (CRC) development through gut microbiota analysis in animal models. Stool status has been associated with potentially adverse gut microbiome profiles from fecal analysis in adults. We examined the association between stool patterns and the prevalence of *pks*⁺ *E. coli* isolated from microbiota in fecal samples of 224 healthy Japanese individuals.

Results: Stool patterns were determined through factorial analysis using a previously validated questionnaire that included stool frequency, volume, color, shape, and odor. Factor scores were classified into tertiles. The prevalence of *pks*⁺ *E. coli* was determined by using specific primers for *pks*⁺ *E. coli* in fecal samples. Plasma and fecal fatty acids were measured via gas chromatography-mass spectrometry. The prevalence of *pks*⁺ *E. coli* was 26.8%. Three stool patterns identified by factorial analysis accounted for 70.1% of all patterns seen (factor 1: lower frequency, darker color, and harder shape; factor 2: higher volume and softer shape; and factor 3: darker color and stronger odor). Multivariable-adjusted odds ratios (95% confidence intervals) of the prevalence of *pks*⁺ *E. coli* for the highest versus the lowest third of the factor 1 score was 3.16 (1.38 to 7.24; *P* for trend = 0.006). This stool pattern exhibited a significant positive correlation with fecal isobutyrate, isovalerate, valerate, and hexanoate but showed a significant negative correlation with plasma eicosenoic acid and α -linoleic acid, as well as fecal propionate and succinate. No other stool patterns were significant.

* Correspondence: cardiovascular0327@mac.com

¹Department of Physical Activity Research, National Institutes of Biomedical Innovation, Health and Nutrition (NIBIOHN), Tokyo 162-8636, Japan

⁸Faculty of Sport Sciences, Waseda University, Saitama 359-1192, Japan

Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Conclusions: These results suggest that stool patterns may be useful in the evaluation of the presence of tumorigenic bacteria and fecal fatty acids through self-monitoring of stool status without the requirement for specialist technology or skill. Furthermore, it may provide valuable insight about effective strategies for the early discovery of CRC.

Keywords: Tumorigenic bacteria, Stool pattern, Fatty acid, Cross-sectional study

Background

Colorectal cancer (CRC) is the second most frequent cause of cancer death and the third most common cancer worldwide. Among all cancers, CRC contributes to a mortality rate of approximately 9% and an incidence of 10% [1]. The majority of disability-adjusted life years lost due to CRC primarily comes from years of life lost (95%), with years lived with disability only contributing 5% [2]. Reducing the number of CRC patients not only has a substantial impact on increased overall longevity for the human race but also has significant effects on reducing mean medical care costs. Thus, the discovery of simple targets that can lead to the early discovery of CRC risk is needed.

The fecal microbiome may contribute to CRC development because it shows changes in the very early stages [3]. Colibactin is a complex secondary metabolite produced by certain *Escherichia coli* strains in the gut that harbor the genomic island (*clb* gene cluster) encoding polyketide synthase (*pks*⁺ *E. coli*) [4–11]. Colibactin has been shown to cause genomic instability to mammalian cells by inducing DNA interstrand cross-links via DNA alkylation [7, 10, 11]. This phenomenon leads to DNA double-strand breaks [6–9] and cell cycle arrest [8]. Therefore, the presence of colibactin-producing *pks*⁺ *E. coli* in the gut microbiome may be a risk factor for CRC and may be a useful target for the identification of groups at high risk for both incidence and progression.

Previous studies have been reported that stool status variables such as stool shape [12–14], frequency [12, 15], and color [16] are associated with gut microbiome profiles from fecal analysis in healthy individuals and patients with acute gastroenteritis. Using a stool pattern approach, which considers a more comprehensive overview of the stool status, can provide more interpretable findings than studying single stool examinations because some stool status variables such as stool frequency and shape are related to each other [12].

We have previously reported that dietary intake was inversely related to the prevalence of *pks*⁺ *E. coli* in Japanese adults [17]. Indeed, dietary intervention has been reported to improve stool status variables, including stool frequency and shape with simultaneous increase in both fecal [18] and plasma [19] fatty acids. Therefore, to further clarify the association between stool status and the prevalence of *pks*⁺ *E. coli*, the association between stool status and fecal and plasma fatty

acids need to be better understood. To our knowledge, these relationships have not been studied. In this study, we aimed: 1) to evaluate the association between stool patterns and prevalence of *pks*⁺ *E. coli* isolated from fecal samples, and 2) to investigate the relationship between stool patterns and plasma and fecal fatty acids in healthy Japanese individuals. We hypothesized that stool pattern variables were associated with the prevalence of *pks*⁺ *E. coli* and the levels of plasma and fecal fatty acids because stool patterns may reflect the gut microbiota.

Results

Participant characteristics

Table 1 shows the characteristics of participants in the analysis cohort. Of the 224 participants included, 60 participants were positive for *pks*⁺ *E. coli* isolated from fecal samples (26.8%). In the comparison between participants who were positive and negative for *pks*⁺ *E. coli*, the presence of *pks*⁺ *E. coli* tended to be lower in women but greater in alcohol drinkers. Other variables were not significantly associated. The minimum detection level of *clb* genes by polymerase chain reaction (PCR) was estimated at a concentration of 10 ng/mL as a DNA template (Fig. 1).

Reproducibility of self-reported stool status

We evaluated the reproducibility of self-reported stool status variables, including stool volume, shape, color, and odor (Table 2). The weighted κ statistics (95% confidence interval [CI]) were 0.96 (0.93 to 0.99) for stool volume, 0.95 (0.92 to 0.98) for stool color, 0.97 (0.95 to 0.99) for stool shape, and 0.92 (0.86 to 0.98) for stool odor. Stool status that was self-reported twice was highly reproducible.

Multivariate analysis for stool patterns and *pks*⁺ *E. coli* carriers

Table 3 shows the stool patterns identified by factorial analysis in this population. Three stool patterns accounted for 70.1% of all patterns seen (factor 1: lower frequency, darker color, and harder shape; factor 2: higher volume and softer shape; and factor 3: darker color and stronger odor).

We evaluated the relationship between the prevalence of *pks*⁺ *E. coli* and stool patterns by multivariate logistic analysis (Table 4). The multivariable-adjusted odds ratios (95% CI) of the prevalence of *pks*⁺ *E. coli* for the

Table 1 Baseline characteristics of participants with or without *pks⁺ E. coli*

	Total (n = 224)	<i>pks⁺ E. coli</i> (n = 60)	<i>pks⁻ E. coli</i> (n = 164)	p-value
Age [years] ^a	58.7 (12.5)	57.2 (13.1)	59.2 (12.2)	0.290
Women [n (%)] ^b	165 (73.7)	37 (61.7)	128 (78.0)	0.014
Body mass index [kg/m ²] ^a	22.4 (2.7)	22.4 (2.6)	22.4 (2.8)	0.859
Current smoking [n (%)] ^b	8 (3.6)	4 (6.7)	4 (2.4)	0.260
Alcohol drinker [n (%)] ^b	116 (51.8)	37 (61.7)	71 (43.3)	0.015
GTC everyday [n (%)] ^b	169 (75.4)	41 (68.3)	128 (78.0)	0.135
Hypertension [n (%)] ^b	39 (17.4)	12 (20.0)	27 (16.5)	0.537
Hyperlipidemia [n (%)] ^b	52 (23.2)	12 (20.0)	40 (24.4)	0.491
FH of cancer [n (%)] ^b	127 (56.7)	29 (48.3)	98 (59.8)	0.127
Antibiotics use [n (%)] ^b	26 (11.7)	8 (13.3)	18 (11.1)	0.648
Probiotics use [n (%)] ^b	28 (12.6)	6 (10.0)	22 (13.6)	0.476
Step counts [step/day] ^a	9610 (3241)	9582 (2887)	9610 (3374)	0.997
Employment [n (%)] ^b	141 (62.9)	43 (71.7)	98 (59.8)	0.102
Sleep time [min/day] ^a	398 (60)	405 (53)	395 (62)	0.251
Hemoglobin [g/dL] ^a	13.4 (1.1)	13.4 (1.3)	13.4 (1.1)	0.701
AST [IU/L] ^a	22.9 (6.9)	22.3 (6.4)	23.1 (7.1)	0.442
ALT [IU/L] ^a	18.3 (10.2)	18.0 (8.5)	18.4 (10.8)	0.794
γ-GTP [IU/L] ^a	29.3 (28.6)	28.7 (25.1)	29.5 (29.9)	0.851
FPG [mg/dL] ^a	87.0 (8.7)	87.7 (9.3)	86.8 (8.4)	0.489
FSI [μU/mL] ^a	4.0 (3.9)	3.6 (1.8)	4.1 (4.5)	0.414
HbA1c [%] ^a	5.5 (0.3)	5.5 (0.4)	5.5 (0.3)	0.527
Triglyceride [mg/dL] ^a	88.7 (57.4)	81.5 (46.2)	91.3 (60.9)	0.259
HDL-C [mg/dL] ^a	68.7 (18.6)	67.7 (21.9)	69.1 (17.3)	0.596
LDL-C [mg/dL] ^a	127 (30)	125 (29)	128 (30)	0.449

ALT alanine transaminase, AST aspartate transaminase, FH family history, FPG fasting plasma glucose, FSI fasting serum insulin, GTC green tea consumption, HbA1c hemoglobin A1c, HDL-C high-density lipoprotein-cholesterol, LDL-C low-density lipoprotein-cholesterol, SD standard deviation, γ-GTP γ-glutamyl transpeptidase.

The p-values shown in bold are statistically significant (p < 0.05).

^aContinuous variables are shown as mean with standard deviation and were analyzed by unpaired t-test. Body mass index was calculated as body weight (kg) divided by height squared (m²)

^bCategorical variables are shown as number of individuals (%) and were analyzed using a X² test

highest versus the lowest third of the factor 1 score was significant at 3.16 (1.38 to 7.24; P for trend = 0.006); however, no significance was observed for any other stool pattern. In addition, stool status such as stool color and shape was significantly associated with the prevalence of *pks⁺ E. coli* (Supplementary Table 1).

Correlation between stool status and plasma and fecal fatty acids

Table 5 shows the association of stool patterns with fatty acids derived from plasma and fecal samples. The factor 1 score was significantly positively correlated with fecal isobutyrate, isovalerate, valerate, and hexanoate but was

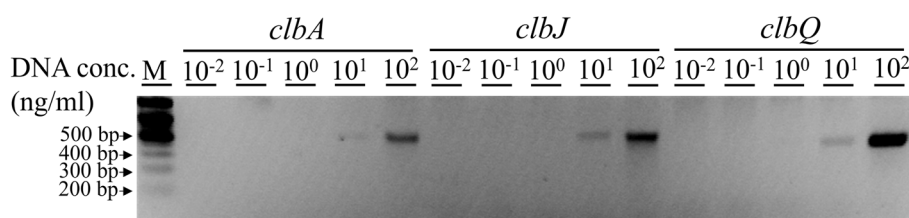


Fig. 1 PCR products amplified at five concentrations (10⁻² to 10² ng/mL) using *clb* genomic DNA purified from a colibactin-producing *E.coli*-50. The *clb* genes and the expected sizes of their amplicons in PCR are as follows: *clbA*, 613 bp; *clbJ*, 544 bp; and *clbQ*, 430 bp. M, DNA marker; conc., concentration

Table 2 Reproducibility of stool status assessed by self-reported questionnaire

	Agreement		Adjacent agreement		Disagreement		Weighted κ statistic	95% CI
	n	(%)	n	(%)	n	(%)		
Stool volume	120	(53.6)	46	(20.5)	17	(7.6)	0.96	(0.93 to 0.99)
Stool color	133	(59.4)	67	(29.9)	7	(3.1)	0.95	(0.92 to 0.98)
Stool shape	161	(71.9)	43	(19.2)	4	(1.8)	0.97	(0.95 to 0.99)
Stool odor	159	(71.0)	62	(27.7)	3	(1.3)	0.92	(0.86 to 0.98)

All values are shown as numbers of people and percentages and weighted κ statistics with 95% confidence interval (CI). The stool status was evaluated by each category as follows: stool volume [1 point (smaller) to 8 points (larger)], stool color [1 point (yellow) to 6 points (dark)], stool shape [1 (severe constipation; separate hard lumps) to 7 (severe diarrhea; liquid consistency with no solid pieces)], and stool odor [1 (no smell) to 4 (severe smell)]. Disagreement was defined as a difference of more than three categories between each variable. The weighted κ statistic can range from 0.00 (perfect disagreement) to 1.00 (perfect agreement)

significantly negatively correlated with plasma eicosenoic acid and α -linoleic acid, as well as fecal propionate and succinate. Other stool patterns showed no significant correlation. The correlations of stool status with plasma and fecal fatty acids, food and beverage consumption, and nutrients intake are tabulated and presented in Supplementary Tables 2, 3, 4 and 5, respectively.

Discussion

In this study, we investigated the relationship between the prevalence of *pks*⁺ *E. coli* and stool patterns through a population-based cohort study. Even after adjustment for confounders, we found a stool pattern (factor 1) that was significantly associated with the prevalence of *pks*⁺ *E. coli*. In addition, this stool pattern was correlated with certain plasma and fecal fatty acids. To the best of our knowledge, this is the first study to show the association of stool patterns with not only the prevalence of *pks*⁺ *E. coli* but also plasma and fecal fatty acids. This association suggests that stool patterns may reflect the presence of tumorigenic bacteria and may provide useful

Table 3 Stool patterns identified by factorial analysis using the 5-stool status

	Factor loadings		
	1	2	3
Stool status			
Stool frequency	-0.817	0.004	0.043
Stool volume	0.006	0.895	-0.035
Stool color	0.534	0.250	0.450
Stool shape	-0.607	0.475	0.012
Stool odor	-0.042	-0.086	0.931
Variance explained by the factor (%)	26.7	21.9	21.5
Cumulative variance (%)	26.7	48.6	70.1

The values are expressed as factor load with the varimax rotation. Bolded values are stable factor load scores (≥ 0.4). The stool status was evaluated by each category as follows: stool frequency [1 point (2 or less times per week) to 6 points (7 or more times per week)], stool volume [1 point (smaller) to 8 points (larger)], stool color [1 point (yellow) to 6 points (dark)], stool shape [1 (severe constipation; separate hard lumps) to 7 (severe diarrhea; liquid consistency with no solid pieces)], and stool odor [1 (no smell) to 4 (severe smell)]

insight about effective strategies for the early discovery of CRC.

Certain risk factors for CRC incidence have been identified in epidemiological studies, including smoking, obesity, diabetes, and high alcohol intake, as well as consumption of red and processed meats [20]. These identified risk factors for CRC [21] and stool status [12–16] have been associated with not only an increased CRC incidence but also potentially adverse gut microbiome profiles. Major research efforts toward the development of stool-based screening tests are ongoing [20], considering that stool status assessment can be conducted through noninvasive self-monitoring without taking time away from routine activities and without the requirement for specialist technology or skill. Recently, the prevalence of *pks*⁺ *E. coli* isolated from the colonic epithelium has been reported to be higher in patients with familial adenomatous polyposis [22], inflammatory bowel disease [23], and CRC [22] than in healthy individuals. Thus, it is important to evaluate the association between comprehensive stool status using a validated tool and the prevalence of tumorigenic bacteria in the gut microbiota.

Our results indicated a significant association between the prevalence of *pks*⁺ *E. coli* and stool pattern. Compared to the sham model, animal models mimicking the natural transmission of colibactin-producing *E. coli* from mothers to neonates have shown lower rates of Firmicutes taxa, Proteobacteria taxa, and microbial species richness, as well as higher DNA repair function [24]. These models have also illustrated an association with gut homeostasis activities, including renewal of the mature epithelium and occurrence of crypt fission [25]. Stool status variables such as shape [12, 13], frequency [12, 15], and color [16] have been associated with higher microbial species richness profiles from fecal analysis in healthy individuals and patients with acute gastroenteritis. Our results support these findings. A previous study showed that the majority of CRC deaths were attributed to non-screening in the United States [26]. Although a causal relationship between the prevalence of *pks*⁺ *E. coli*

Table 4 Odds ratios for stool patterns and prevalence of *pks*⁺ *E. coli* carriers

Stool patterns	Number or ORs (95% CI)						<i>p</i> for trend ^a
	T1		T2		T3		
Factor 1							
<i>n</i>	77		71		76		
Case [<i>n</i> (%)]	14	(18.2)	16	(22.5)	30	(39.5)	
Model 1 ^b	1.00	(Ref)	1.42	(0.61 to 3.27)	3.62	(1.63 to 8.03)	0.002
Model 2 ^c	1.00	(Ref)	1.12	(0.47 to 2.67)	3.16	(1.38 to 7.24)	0.006
Factor 2							
<i>n</i>	75		74		75		
Case [<i>n</i> (%)]	17	(22.7)	22	(29.7)	21	(28.0)	
Model 1 ^b	1.00	(Ref)	1.42	(0.67 to 2.98)	1.12	(0.52 to 2.40)	0.673
Model 2 ^c	1.00	(Ref)	1.46	(0.67 to 3.18)	1.09	(0.49 to 2.44)	0.767
Factor 3							
<i>n</i>	75		75		74		
Case [<i>n</i> (%)]	18	(24.0)	22	(29.3)	20	(27.0)	
Model 1 ^b	1.00	(Ref)	1.10	(0.51 to 2.35)	1.05	(0.49 to 2.23)	0.403
Model 2 ^c	1.00	(Ref)	1.18	(0.53 to 2.62)	1.12	(0.50 to 2.51)	0.301

Ref reference. The prevalence rates of *pks*⁺ *E. coli* are shown as numbers of people and percentages. The detail of the three stool patterns are as follows. Factor 1: lower frequency, darker color, and harder shape; factor 2: higher volume and softer shape; and factor 3: darker color and stronger odor

^aStatistical analysis was carried out using the likelihood ratio test for multivariate logistic analysis, and the odds ratios (ORs) and 95% confidence intervals (CI) were estimated. Bold *p* values are statistically significant (*p* < 0.05)

^bModel 1 was adjusted for age (continuous) and sex (male or female)

^cModel 2 was as model 1 plus mutual adjustment for body mass index (continuous), family history of cancer (yes or no), smoking status (never smoker, past smoker, or current smoker), step counts (continuous), alcohol drinker (yes or no), and green tea consumption (continuous)

as a tumorigenic bacteria and an increased risk of CRC has not been established well, it may be potential benefits of evaluating the presence of *pks*⁺ *E. coli* as a target for early prognostication in populations with a high risk of CRC. Our results suggest that stool patterns might be markers associated with the prevalence of tumorigenic bacteria in healthy individuals. Longitudinal objective monitoring of a person's stool status from serial samples obtained from an individual's excreta at home, as previously suggested [27], may be the most reasonable and cost-effective method for the early detection of risk factors for CRC.

The stool pattern associated with the presence of *pks*⁺ *E. coli* (i.e., factor 1) exhibited a significant negative correlation with the consumption of noodles, alcoholic and non-alcoholic beverages, fruit and vegetable juices, green tea, niacin, and sodium (Supplementary Tables 4 and 5). We had previously shown that green tea consumption was negatively associated with the presence of *pks*⁺ *E. coli* [17], and this previous finding supports our present results. A previous study using fecal samples collected from healthy adults reported that green tea consumption reduces the microbial functional pathways' abundance relevance to carcinogenesis after 2 weeks of intervention [28]. These results suggest that green tea consumption significantly reduces the prevalence of *pks*⁺ *E. coli*. Via the suppression of the growth of certain microorganisms

in the gut microbiome. Taking information from previous studies, we speculate that gut microbiota and dietary components interacted to generate biologically active molecules, which influenced gut secretion and motility, and that this could play a fundamental role in stool status [29, 30] and affect the prevalence of *pks*⁺ *E. coli*. Nonetheless, our results did not substantially change even after adjustment for green tea consumption, indicating that its effect might be weak. Hence, a well-designed study that further evaluates dietary intake, microbiota, and biologically active molecules in the gut in detail is required.

Nutrients derived from ingested food are utilized by the gut microbiome, with certain preferred energy sources such as short-chain fatty acids (SCFAs) for colonocytes [31, 32]. These metabolites can suppress inflammation and carcinogenesis via their effects on immunity, gene expression, and epigenetic modulation [29–32]. Some plasma fatty acids have been shown to be inversely or positively associated with the presence of colon adenomas [33] and an increased risk of CRC in middle-aged adults [34]. Furthermore, studies involving CRC patients have reported lower propionate and butyrate levels [35] and higher valeric acid, isobutyric acid, and isovaleric acid levels [36] in SCFAs derived from fecal samples, as compared to healthy controls. SCFA production is reduced in patients with diarrhea, as compared to those

Table 5 Correlation between stool patterns and plasma and fecal fatty acids

	Structure	n	Unit	Median	Interquartile range	Stool status pattern		
						1	2	3
Plasma								
Lauric acid	C12:0	221	µg/ml	1.8	(1.2 to 2.7)	0.02	-0.05	-0.05
Myristic acid	C14:0	223	µg/ml	21.4	(16.4 to 28.8)	-0.09	0.03	-0.07
Palmitic acid	C16:0	223	µg/ml	668.7	(594.6 to 749.7)	-0.07	0.06	-0.04
Stearic acid	C18:0	223	µg/ml	227.3	(203.3 to 259.3)	-0.07	0.07	-0.03
Arachidic acid	C20:0	223	µg/ml	8.9	(7.8 to 10.0)	-0.11	-0.06	-0.05
Behenic acid	C22:0	223	µg/ml	21.7	(18.9 to 24.9)	0.00	-0.02	-0.03
Lignoceric acid	C24:0	223	µg/ml	19.4	(0.8 to 22.8)	0.03	-0.02	-0.02
Total saturated fatty acids		223	µg/ml	969.9	(862.4 to 1092.5)	-0.07	0.06	-0.04
Oleic acid	C18:1n-9	223	µg/ml	568.6	(485.6 to 659.0)	-0.11	0.10	-0.01
Eicosenoic acid	C20:1n-9	223	µg/ml	4.2	(3.5 to 5.3)	-0.16	*	0.06
5,8,11-Eicosatrienoic acid	C20:3n-9	217	µg/ml	2.4	(1.5 to 3.2)	-0.05	-0.03	0.07
Erucic acid	C22:1n-9	48	µg/ml	1.2	(1.1 to 1.3)	-0.04	-0.04	-0.12
Nervonic acid	C24:1n-9	223	µg/ml	36.1	(1.6 to 42.2)	-0.03	0.04	-0.10
Total n-9 fatty acids		223	µg/ml	610.7	(514.5 to 701.9)	-0.12	0.10	-0.01
Palmitoleic acid	C16:1n-7	223	µg/ml	48.7	(38.1 to 60.7)	-0.10	0.04	-0.08
Linoleic acid	C18:2n-6	223	µg/ml	956.8	(846.0 to 1111.3)	-0.02	0.05	-0.02
γ-linolenic acid	C18:3n-6	223	µg/ml	8.9	(5.8 to 12.9)	-0.06	-0.03	0.04
Eicosadienoic acid	C20:2n-6	223	µg/ml	7.2	(6.1 to 8.4)	-0.10	0.01	0.02
Dihomo-gamm-linolenic acid	C20:3n-6	223	µg/ml	38.7	(29.8 to 46.1)	-0.07	-0.05	0.05
Arachidonic acid	C20:4n-6	223	µg/ml	214.8	(184.5 to 249.7)	0.03	0.05	-0.01
Docosatetraenoic acid	C22:4n-6	223	µg/ml	4.8	(3.9 to 5.8)	0.04	0.03	0.06
Total n-6 fatty acids		223	µg/ml	1234.0	(1104.4 to 1389.5)	-0.01	0.03	-0.01
Myristoleic acid	C14:1n-5	221	µg/ml	1.2	(0.8 to 1.8)	-0.08	0.02	-0.10
α-linoleic acid	C18:3n-3	223	µg/ml	22.3	(16.7 to 30.0)	-0.14	*	0.03
Eicosapentaenoic acid	C20:5n-3	223	µg/ml	65.9	(43.5 to 98.6)	-0.12	-0.01	-0.12
Docosapentaenoic acid	C22:5n-3	223	µg/ml	16.5	(0.8 to 21.7)	-0.03	0.02	-0.05
Docosahexaenoic acid	C22:6n-3	223	µg/ml	137.4	(6.4 to 173.7)	-0.04	0.03	-0.10
Total n-3 fatty acids		223	µg/ml	236.4	(152.2 to 301.9)	-0.10	0.03	-0.11
Mono fatty acids		223	µg/ml	660.3	(560.4 to 756.6)	-0.12	0.10	-0.03
Polyunsaturated fatty acid		223	µg/ml	1482.0	(1292.1 to 1667.6)	-0.05	0.05	-0.07
n-3/n-6 fatty acids		223		0.18	(0.13 to 0.24)	-0.10	-0.01	-0.12
Total fatty acids		223	µg/ml	3093.9	(2749.6 to 3512.7)	-0.07	0.07	-0.05
Fecal								
Formate	C1:0	58	mol/g	2.66	(0.83 to 4.59)	0.17	-0.10	-0.07
Acetate	C2:0	161	mol/g	57.41	(44.66 to 67.12)	-0.06	0.09	0.06
Propionate	C3:0	161	mol/g	16.40	(11.59 to 23.96)	-0.30	**	0.15
Lactate	C3:0	19	mol/g	0.08	(0.03 to 0.30)	0.04		-0.08
Isobutyrate	C4:0	161	mol/g	1.57	(0.95 to 2.19)	0.21	**	-0.11
Butyrate	C4:0	161	mol/g	10.30	(6.59 to 17.52)	-0.07		0.06
Succinate	C4:0	143	mol/g	0.18	(0.09 to 0.34)	-0.21	**	0.05
Isovalerate	C5:0	161	mol/g	1.01	(0.57 to 1.50)	0.27	**	-0.12

Table 5 Correlation between stool patterns and plasma and fecal fatty acids (Continued)

	Structure	n	Unit	Median	Interquartile range	Stool status pattern			
						1	2	3	
Valerate (Pentanoate)	C5:0	160	mol/g	1.36	(0.66 to 2.06)	0.15	*	-0.03	-0.08
Hexanoate	C6:0	161	mol/g	0.06	(0.02 to 0.28)	0.32	**	-0.07	0.06
Total SCFA		161	mol/g	97.02	(72.12 to 116.97)	-0.13		0.11	0.03

Statistical analysis was performed by Spearman's correlation analysis. The statistical significance p value is indicated as follows: if $p < 0.05$, single asterisk (*); if $p < 0.01$ double asterisk (**). If the results presented a positive correlation, the participants with higher adherence to each stool patterns mean to relate higher plasma and fecal fatty acids (conversely, a negative correlation indicates that they mean lower it)

without diarrhea [37]. Additionally, inhibition of SCFA synthesis via the administration of polyethylglycol and antibiotics results in diarrhea [38]. The distal colon transit, as reflected in stool frequency, is associated with not only plasma acetate and fecal SCFAs [39] but also microbiota diversity, especially the Firmicutes taxa (*Faecalibacterium*, *Lactococcus*, and *Roseburia*) [40]. Previous studies reported that certain plasma fatty acids, including α -linoleic acid [41] and certain fecal SCFAs such as propionate [35] and isovaleric acid [36], were associated with a higher CRC incidence. Our results indicated that the stool pattern showing a relationship with the prevalence of *pks*⁺ *E. coli* was also significantly correlated these fatty acids; thus, the findings of these previous studies support our results. While detailed mechanisms and causal relationships should be clarified in further studies, we can conclude that fecal matter is not just a simple waste material but may be useful in the evaluation of the presence of tumorigenic bacteria and fecal fatty acids by stool status via comprehensive examination of variables such as color, shape, frequency, volume, and odor.

The strength of this study is in finding a verified association between stool patterns and plasma and fecal fatty acids. The multifaceted, self-reported questionnaires used for stool status assessment had previously been validated against objective fecal characteristics as well [42]. In addition, we showed that twice self-reported stool status was highly reproducible and believe that it is unlikely for there to have been misclassification when done in this manner. Thus, this study might generate a new hypothesis for the association between the prevalence of *pks*⁺ *E. coli* as tumorigenic bacteria and stool pattern. However, this study has a number of methodological limitations. First, even if the effects of confounders was minimized using multivariate analysis with adjustment for known covariates, being a cross-sectional study, the present study is unable to theorize about the temporal and direct causality of the observed association between stool patterns and the prevalence of *pks*⁺ *E. coli*. Second, this study detected the *clb* gene cluster in the DNA extracted from fecal samples, not in the DNA isolated from the colonic epithelium. A previous study evaluated the prevalence of *pks*⁺ *E. coli* using the selective

cultivation method [23]. Our results indicated that the *clb* genes (i.e., *clbA*, *clbJ*, and *clbQ*) were detected as PCR products at a concentration of more than 10 ng/mL of the DNA template. The prevalence of *pks*⁺ *E. coli* might be underestimated because it is considered that the *clb* genes could not be detected at concentrations lower than 10 ng/mL in the fecal samples. However, our previous study [17] showed that the prevalence of *pks*⁺ *E. coli* isolated from fecal matter was relatively similar to that in previous reports investigating the prevalence of *pks*⁺ *E. coli* using the selective cultivation method [23]. Therefore, evaluating the concordance rate of the prevalence of *pks*⁺ *E. coli*, defined using these two different methods for the same subject, will be necessary. Third, although our results showed that a softer stool shape was negatively associated with the prevalence of *pks*⁺ *E. coli*, it is unclear whether the prevalence of *pks*⁺ *E. coli* was lower in participants with diarrhea who had softer stools or whether the reverse was true. In addition, we were unable to completely exclude systematic error due to self-reporting, and we could not account for unmeasured confounding factors associated with stool status in this observational study. For example, stool color is mainly characterized by stercobilin (urobilin), an orange pigment and an oxidized metabolite of urobilinogen [43]. Stercobilin derived from bile pigment is responsible for the brown color of human feces. As we did not directly measure stercobilin and bile acids in all participants, we could not account for their possible effects on the results; nonetheless, our results were similar after adjustment for bile acids in a subsection of participants with available bile acid data. It is necessary to further verify our results with further studies including patients and community-dwelling residents with symptoms such as diarrhea and constipation. Finally, there is the possibility of sampling bias due to the more health-aware nature of the participants in this study than in the general population. Of 750 participants in the Nutrition and Exercise Intervention Study (NEXIS) cohort study, 259 adults agreed to participate. As the participation rate was relatively low, selection bias might have occurred. Additionally, the participants were all living in the Tokyo metropolitan area in Japan and the mean age was 58 years. These limitations may prevent the generalization

of our results. Therefore, prospective cohort studies with larger randomized samples should be conducted to further investigate the association between the prevalence of *pks*⁺ *E. coli* and stool patterns.

Conclusion

These results suggest that an adverse stool pattern is positively associated with the prevalence of *pks*⁺ *E. coli*. Given the rapidly increasing incidence and mortality rates of CRC worldwide, its early discovery is important to both enable people to stay healthy and limit the burden on healthcare-related costs. Therefore, stool patterns may be useful in the evaluation of the presence of tumorigenic bacteria and fecal fatty acids through self-monitoring of stool status without the requirement for specialist technology or skill. Furthermore, it may provide valuable insight about effective strategies for the early discovery of CRC.

Methods

Participants and study procedure

This cross-sectional study utilized data from the NEXIS cohort study [17]. This cohort study has been managed by the National Institutes of Biomedical Innovation, Health and Nutrition (NIBIOHN) since 2012 and aims to evaluate the association between gut microbiota and lifestyle, including dietary intake and physical activity (NIBIOHN: no. kenei 102; clinical trial registration number: NCT00926744). Of 750 participants in the NEXIS cohort study, 259 adults aged 27–79 years who were living in the Tokyo metropolitan area in Japan agreed to participate in this study (NIBIOHN: no. kenei 3-04; clinical trial registration number: UMIN000023270). This study was approved by the ethics review board of the Research Ethical Review Committee of the NIBIOHN. After the study procedures and risks associated with the participation in this study were explained, written informed consent was obtained from all participants before data acquisition. This study was conducted in accordance with the principles of the Declaration of Helsinki.

A kit for fecal collection and storage and the questionnaire for the lifestyle survey were mailed to the participants. They were instructed to complete the questionnaire to record pertinent lifestyle variables (e.g. medical history, smoking habit, dietary intake, and stool status) and to collect fecal samples approximately 7 mm in diameter (soybean size) at home. Dietary intake was evaluated using a previously validated brief-type self-administered diet history questionnaire [44]. To measure daily step counts as an objective form of physical activity, we used a triaxial accelerometer (Actimarker EW4800; Panasonic Co., Ltd., Japan). The participants were instructed to bring their fecal samples and questionnaires to the NIBIOHN within

a week after answering the questionnaires and finishing the serial fecal collection. Subsequently, they underwent physical and health examinations such as anthropometry and blood tests in the NIBIOHN. Investigators, registered dietitians, or nurses checked the questionnaires and interviewed those with unclear responses or unanswered questions to confirm answers. Blood samples were used as a biochemical examination for conventional risk factors for lifestyle-related diseases, with close attention placed on variables such as low-density lipoprotein-cholesterol, hemoglobin A1c, and triglycerides. The collected feces, serum, and plasma were immediately placed in a sealed container and stored as individual sample types to avoid cross-contamination between samples in a –20 °C freezer.

Of the participants initially included in this study ($n = 259$), we excluded those with diabetes mellitus ($n = 13$), history of cancer ($n = 12$), cardiovascular disease ($n = 6$), gastrointestinal disease ($n = 3$), and renal failure ($n = 1$). Ultimately, 224 participants were included in the final analysis.

Confirmation of *pks*⁺ *E. coli* by PCR

Bacterial genomic DNA was extracted from frozen fecal samples. Details of this protocol have been reported elsewhere [17, 45, 46]. To confirm the presence of *pks*⁺ *E. coli*, we performed PCR to amplify the genes from the *clb* cluster using bacterial genomic DNA as a template at a concentration of approximately 10 µg/mL. PCR was conducted using SapphireAmp Fast PCR Master Mix (Takara Bio Inc., Shiga, Japan) according to the manufacturer's protocol. The PCR conditions were as follows: (1) 94 °C for 2 min; (2) 98 °C for 5 s; (3) 63 °C for 5 s; (4) 72 °C for 20 s; (5) repeat (2) to (4) for 30 cycles. The primers used in the PCR experiments were as follows: *clbB* forward primer: 5'-tggtccgtttgtgtggttcagcg-3', reverse primer: 5'-gtgcgctgaccattgaagattccg-3'; *clbJ* forward primer: 5'-tggcctgtattgaaagaccggtt-3', reverse primer: 5'-aatgggaacggttgatgacgatgct-3'; *clbQ* forward primer: 5'-ctgtgtcttacgatggtggatgccg-3', reverse primer: 5'-gcattaccagattgcagcatgcc-3'. The *clb* genes and the expected sizes of their amplicons in PCR were as follows: *clbA*, 613 bp; *clbJ*, 544 bp; and *clbQ*, 430 bp. The amplified DNA was electrophoresed on 3% agarose gel (100 V, 5 min) using the Tris-acetate-EDTA buffer and visualized by ethidium bromide staining. In this analysis, samples amplified with the appropriate amplicon length in the three *clb* genes were defined as *pks*⁺ *E. coli*-positive individuals [17]. Furthermore, in order to determine the minimum detection level of *clb* genes using the PCR method, we verified the genomic DNA purified from a colibactin-producing *E. coli*-50 at each concentration of 0.01–100 ng/mL using a PCR template [45].

Evaluation of stool status

Stool status was assessed using the multifaced self-reported questionnaire called the “intestinal visible sheet”, which covers the 5-stool status variables (frequency, volume, color, shape, and odor) and was previously developed and validated against objective measurements of fecal characteristics including fecal weight, moisture, hardness, and color in adults [42]. In the NEXIS cohort study, all participants were given similar stool questionnaires assessing both “habitual stool status” and “stool status when collecting fecal samples (excluding stool frequency)” at the same time. We evaluated the reproducibility of the results by comparing these variables because variables evaluated based on self-reporting questionnaires may be affected by recall bias. We used the habitual stool status data in all analysis because it simultaneously evaluated all five stool statuses.

Measurement of plasma fatty acids

To investigate plasma and fecal fatty acids, we used the frozen stored plasma and fecal samples. Total lipids were extracted from 0.4 mL of plasma following the methodology reported by Folch et al. [47]. After hydrolysis with 5% KOH, fatty acids were extracted with hexane and tricosanoic acid (C23:0) as the internal standard. Methyl-esterified fatty acids were prepared with a trimethylsilylating reagent and subjected to gas chromatography (GC). GC-electrospray ionization mass spectrometry (ESI/MS) analysis was performed using Shimadzu GC-2010 (Shimadzu Corporation, Kyoto, Japan) equipped with a hydrogen flame ionization detector. A glass column (40 m × 0.3 mm in volume) was coated with diethylene glycol succinate. An AOC-20i autoinjector (Shimadzu Corporation, Kyoto, Japan) was employed for sample injection. Nitrogen gas was used as a carrier gas and delivered at a flow rate of 25 mL min⁻¹. Injection volume was set at 20 µL with a split ratio of 80:1. Column temperature was maintained at 180 °C, and the injection port and detector cell were kept at 350 °C and 250 °C, respectively. The 24 types of plasma fatty acids with chain lengths comprising 12–24 carbons were measured through this analysis.

Measurement of fecal short-chain fatty acids

Furthermore, 5–10 mg of feces was mixed with 90 µL of Milli-Q and 10 µL of 2 mM internal standard containing acetic acid, butyric acid, and crotonic acid for 5 min. The mixture was homogenized with 50 µL of 36% HCl and 200 µL of 97% diethyl ether and was centrifuged at 3000 rpm for 10 min at room temperature. Subsequently, 80 µL of the supernatant organic layer was transferred to a new glass vial and combined with 16 µL of N-tert-butyltrimethylsilyl-N-methyltrifluoroacetamide as a derivatization reagent. The vials were immediately capped

tightly with electronic crimper (Agilent), incubated for 20 min in an 80 °C water bath, and then left at room temperature in the dark for 48 h for derivatization. The derivatized samples were analyzed using a GC-MS-TQ8040 gas chromatograph mass spectrometer (Shimadzu Corporation, Kyoto, Japan), and the injection was performed using an AOC-20i autoinjector (Shimadzu Corporation, Kyoto, Japan). The capillary column was a BPX5 column (0.25 mm × 30 m × 0.25 µm; Shimadzu GLC). Pure helium gas was used as a carrier gas and delivered at a flow rate of 1.2 mL min⁻¹. The head pressure was 72.8 kPa with split (split ratio of 30:1). The injection port and interface temperatures were 230 °C and 260 °C, respectively. This analysis measured the 10 types of fecal SCFAs (C1:0–C6:0).

Statistical analysis

Characteristics of participants with and without *pks*⁺ *E. coli* were compared using the baseline characteristics used in a previous study [17]. Continuous variables were presented as means and standard deviations, with differences between the two groups evaluated using the unpaired t-test. Categorical variables were expressed as numbers and percentages, with differences between the two groups evaluated using the chi-square test.

The agreement, adjacent agreement, and disagreement from the twice-evaluated stool status variables were expressed as numbers and percentages. Disagreement was defined as a difference of more than three categories between each variable. In order to evaluate the reproducibility of the variables in self-reported stool status, we used a weighted κ statistic with 95% CI [48].

To identify the primary stool patterns, we used factorial analysis with varimax rotation (orthogonal transformation) to derive non-correlated factors [49]. This approach maintained a greater interpretability because each factor could be noted independent of the others with distribution explained by the variance among the individual components. We considered the scree plot and eigenvalues to determine the number of factors to retain by minimizing the number of indicators that had high loading on one factor [49]. For these reasons, we identified three stool patterns of interest from the five stool statuses. We considered stable factor load to be scored greater than 0.4 [50]. For every participant, we calculated factor scores for each of the three retained factors by summing the scores multiplied by factor load across all stool statuses. Participants were classified into tertiles (Ts) of factor score for each stool pattern.

The prevalence of *pks*⁺ *E. coli* in each tertile for stool patterns was shown as number of cases and percentage. To adjust for confounders between the prevalence of *pks*⁺ *E. coli* and the stool patterns, we used the likelihood

ratio test for multivariate logistic analysis including baseline covariates. The odds ratio (ORs) and 95% CI were estimated. These analyses were verified on two models as follows: Model 1 was adjusted for age (continuous) and sex (female or male); Model 2 was as Model 1 plus mutual adjustment by body mass index (continuous), family history of cancer (yes or no), smoking status (never smoker, past smoker, or current smoker), step count (continuous), alcohol drinker (yes or no), and green tea consumption (continuous). These variables were decided in accordance with covariates used in our previous study [17]. We calculated the prevalence OR of *pks*⁺ *E. coli* with the lowest tertile of each stool pattern as reference. Plasma and fecal fatty acid measurements were expressed as median and interquartile values. Instances wherein fatty acids could not be detected during GC/MS analysis were handled as missing data. In order to evaluate the association of plasma and fecal fatty acids as well as dietary intake with stool patterns, we used Spearman's correlation coefficient.

For statistical significance, the *p* value was set to < 0.05 (double-sided). All statistical analyses were performed using JMP Pro for Windows (SAS Institute, Inc., Cary, NC, USA).

Abbreviations

ALT: Alanine transaminase; AST: Aspartate transaminase; CI: Confidence Interval; *clb*: Colibactin; CRC: Colorectal cancer; ESI/MS: Electrospray Ionization Mass Spectrometry; FH: Family History; FPG: Fasting Plasma Glucose; FSI: Fasting Serum Insulin; GC: Gas Chromatography; GTC: Green Tea Consumption; HbA1c: Hemoglobin A1c; HDL-C: High-Density Lipoprotein-Cholesterol; LDL-C: Low-Density Lipoprotein-Cholesterol; NEXIS: Nutrition and Exercise Intervention Study; NIBIOHN: National Institutes of Biomedical Innovation, Health and Nutrition; OR: Odds Ratio; PCR: Polymerase Chain Reaction; *pks*⁺ *E. coli*: *Escherichia coli* containing polyketide synthase; Ref: Reference; SCFAs: Short-Chain Fatty Acids; SD: Standard Deviation; Ts: Tertiles; γ -GTP: γ -Glutamyl transpeptidase

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12866-021-02255-6>.

Additional file 1: Table S1. Odds ratios for stool status and prevalence of *pks*⁺ *E. coli* carrier calculated by multivariate logistic regression analysis. **Table S2.** Correlation between stool status and plasma and fecal fatty acids. **Table S3.** Correlation between fecal short-chain fatty acids and plasma fatty acids. **Table S4.** Correlation between stool patterns and food and beverage consumption. **Table S5.** Correlation between stool patterns and nutrients intake

Acknowledgements

We are grateful to all of the participants who provided fecal samples for use in this research and to the members of the Physical Activity Research Laboratory at the National Institute of Health and Nutrition, NIBIOHN.

Authors' contributions

The authors' responsibilities were as follows. DW and MM wrote the initial draft of the manuscript. MM, HM, and KW designed the study and interpretation of data and assisted in preparation of the manuscript. HM, HO, KK, KT, JK, and MM contributed to data collection. YT, MS, YO, NM, KW, and KW contributed to the analysis of the evaluation of the *pks*⁺ *E. coli* carrier. KT-M, NK, and TK measured fatty acids in plasma and stool. All authors

critically reviewed the manuscript and approved the final version, and agreed to be accountable for all aspects of the work and ensure that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Funding

This study was supported by the Grant-in-Aid for Scientific Research C (JP16K00944), the Health and Labor Sciences Research Grant (201709002B) to M. Miyachi, and the Development of Innovative Research on Cancer Therapeutics from Japan Agency for Medical Research and Development (16ck0106243h0001) to K. Watanabe. These funding were used for the purpose of designing the study, and for data collection and analysis.

Availability of data and materials

The datasets used during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

This study was approved by the ethics review board of the Research Ethical Review Committee of the National Institutes of Biomedical Innovation, Health and Nutrition. After the study procedures and risks associated with the participation in this study were explained, written informed consent was obtained from all participants before data acquisition.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Department of Physical Activity Research, National Institutes of Biomedical Innovation, Health and Nutrition (NIBIOHN), Tokyo 162-8636, Japan.

²Department of Clinical Nutrition, National Institutes of Biomedical Innovation, Health and Nutrition (NIBIOHN), Tokyo 162-8636, Japan.

³Department of Pharmaceutical Sciences, University of Shizuoka, Shizuoka 422-8526, Japan. ⁴School of Food and Nutritional Sciences, University of Shizuoka, Shizuoka 422-8526, Japan. ⁵Laboratory of Vaccine Materials, Center for Vaccine and Adjuvant Research, and Laboratory of Gut Environmental System, National Institutes of Biomedical Innovation, Health and Nutrition (NIBIOHN), Osaka 567-0085, Japan. ⁶Intestinal Microbiota Project, Kanagawa Institute of Industrial Science and Technology, Kanagawa 243-0435, Japan. ⁷Division of Diabetes and Metabolism, The Institute for Medical Science, Asahi Life Foundation, Tokyo 103-0002, Japan. ⁸Faculty of Sport Sciences, Waseda University, Saitama 359-1192, Japan.

Received: 1 November 2020 Accepted: 9 June 2021

Published online: 28 June 2021

References

1. Ferlay J, Colombet M, Soerjomataram I, Mathers C, Parkin DM, Pineros M, et al. Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. *Int J Cancer*. 2019;144(8):1941–53. <https://doi.org/10.1002/ijc.31937>.
2. Global Burden of Disease Cancer C, Fitzmaurice C, Abate D, Abbasi N, Abbastabar H, Abd-Allah F, et al. Global, Regional, and National Cancer Incidence, Mortality, Years of Life Lost, Years Lived With Disability, and Disability-Adjusted Life-Years for 29 Cancer Groups, 1990 to 2017: A systematic analysis for the global burden of disease study. *JAMA Oncol*. 2019. <https://doi.org/10.1001/jamaoncol.2019.2996>.
3. Yachida S, Mizutani S, Shiroma H, Shiba S, Nakajima T, Sakamoto T, et al. Metagenomic and metabolomic analyses reveal distinct stage-specific phenotypes of the gut microbiota in colorectal cancer. *Nat Med*. 2019;25(6):968–76. <https://doi.org/10.1038/s41591-019-0458-7>.
4. Healy AR, Wernke KM, Kim CS, Lees NR, Crawford JM, Herzon SB. Synthesis and reactivity of precolibactin 886. *Nat Chem*. 2019;11(10):890–8. <https://doi.org/10.1038/s41557-019-0338-2>.
5. Li ZR, Li J, Cai W, Lai JYH, McKinnie SMK, Zhang WP, et al. Macrocyclic colibactin induces DNA double-strand breaks via copper-mediated oxidative

- cleavage. *Nat Chem*. 2019;11(10):880–9. <https://doi.org/10.1038/s41557-019-0317-7>.
6. Dziubanska-Kusibab PJ, Berger H, Battistini F, Bouwman BAM, Iftekhar A, Katainen R, et al. Colibactin DNA-damage signature indicates mutational impact in colorectal cancer. *Nat Med*. 2020;26(7):1063–9. <https://doi.org/10.1038/s41591-020-0908-2>.
 7. Pleguezuelos-Manzano C, Puschhof J, Rosendahl Huber A, van Hoeck A, Wood HM, Nomburg J, et al. Mutational signature in colorectal cancer caused by genotoxic pks(+) *E. coli*. *Nature*. 2020;580(7802):269–73. <https://doi.org/10.1038/s41586-020-2080-8>.
 8. Cuevas-Ramos G, Petit CR, Marcq I, Boury M, Oswald E, Nougayrede JP. *Escherichia coli* induces DNA damage in vivo and triggers genomic instability in mammalian cells. *Proc Natl Acad Sci U S A*. 2010;107(25):11537–42. <https://doi.org/10.1073/pnas.1001261107>.
 9. Nougayrede JP, Homburg S, Taieb F, Boury M, Brzuszkiewicz E, Gottschalk G, et al. *Escherichia coli* induces DNA double-strand breaks in eukaryotic cells. *Science*. 2006;313(5788):848–51. <https://doi.org/10.1126/science.1127059>.
 10. Wilson MR, Jiang Y, Villalta PW, Stornetta A, Boudreau PD, Carra A, et al. The human gut bacterial genotoxin colibactin alkylates DNA. *Science*. 2019;363(6428). <https://doi.org/10.1126/science.aar7785>.
 11. Xue M, Kim CS, Healy AR, Wernke KM, Wang Z, Frischling MC, et al. Structure elucidation of colibactin and its DNA cross-links. *Science*. 2019;365(6457). <https://doi.org/10.1126/science.aax2685>.
 12. Hadizadeh F, Walter S, Belheouane M, Bonfiglio F, Heinsen FA, Andreasson A, et al. Stool frequency is associated with gut microbiota composition. *Gut*. 2017;66(3):559–60. <https://doi.org/10.1136/gutjnl-2016-311935>.
 13. Vandeputte D, Falony G, Vieira-Silva S, Tito RY, Joossens M, Raes J. Stool consistency is strongly associated with gut microbiota richness and composition, enterotypes and bacterial growth rates. *Gut*. 2016;65(1):57–62. <https://doi.org/10.1136/gutjnl-2015-309618>.
 14. Takagi T, Naito Y, Inoue R, Kashiwagi S, Uchiyama K, Mizushima K, et al. Differences in gut microbiota associated with age, sex, and stool consistency in healthy Japanese subjects. *J Gastroenterol*. 2019;54(1):53–63. <https://doi.org/10.1007/s00535-018-1488-5>.
 15. Kwon HJ, Lim JH, Kang D, Lim S, Park SJ, Kim JH. Is stool frequency associated with the richness and community composition of gut microbiota? *Intest Res*. 2019;17(3):419–26. <https://doi.org/10.5217/ir.2018.00149>.
 16. Castano-Rodriguez N, Underwood AP, Merif J, Riordan SM, Rawlinson WD, Mitchell HM, et al. Gut Microbiome Analysis Identifies Potential Etiological Factors in Acute Gastroenteritis. *Infect Immun*. 2018;86(7). <https://doi.org/10.1128/IAI.00060-18>.
 17. Watanabe D, Murakami H, Ohno H, Tanisawa K, Konishi K, Tsunematsu Y, et al. Association between dietary intake and the prevalence of tumorigenic bacteria in the gut microbiota of middle-aged Japanese adults. *Sci Rep*. 2020;10(1):15221. <https://doi.org/10.1038/s41598-020-72245-7>.
 18. Venancio VP, Kim H, Sirven MA, Tekwe CD, Honvoh G, Talcott ST, et al. Polyphenol-rich Mango (*Mangifera indica* L.) Ameliorate Functional Constipation Symptoms in Humans beyond Equivalent Amount of Fiber. *Mol Nutr Food Res*. 2018;62(12):e1701034. <https://doi.org/10.1002/mnfr.201701034>.
 19. Jakobsen LH, Wirth R, Smoliner C, Klebach M, Hofman Z, Kondrup J. Gastrointestinal tolerance and plasma status of carotenoids, EPA and DHA with a fiber-enriched tube feed in hospitalized patients initiated on tube nutrition: randomized controlled trial. *Clin Nutr*. 2017;36(2):380–8. <https://doi.org/10.1016/j.clnu.2016.02.001>.
 20. Brenner H, Kloor M, Pox CP. Colorectal cancer. *Lancet*. 2014;383(9927):1490–502. [https://doi.org/10.1016/S0140-6736\(13\)61649-9](https://doi.org/10.1016/S0140-6736(13)61649-9).
 21. Brennan CA, Garrett WS. Gut microbiota, inflammation, and colorectal Cancer. *Annu Rev Microbiol*. 2016;70:395–411. <https://doi.org/10.1146/annurev-micro-102215-095513>.
 22. Dejea CM, Fathi P, Craig JM, Boleij A, Taddese R, Geis AL, et al. Patients with familial adenomatous polyposis harbor colonic biofilms containing tumorigenic bacteria. *Science*. 2018;359(6375):592–7. <https://doi.org/10.1126/science.aah3648>.
 23. Arthur JC, Perez-Chanona E, Muhlbauer M, Tomkovich S, Uronis JM, Fan TJ, et al. Intestinal inflammation targets cancer-inducing activity of the microbiota. *Science*. 2012;338(6103):120–3. <https://doi.org/10.1126/science.1224820>.
 24. Tronnet S, Floch P, Lucarelli L, Gaillard D, Martin P, Serino M, et al. The Genotoxin Colibactin Shapes Gut Microbiota in Mice, mSphere. 2020;5(4). <https://doi.org/10.1128/mSphere.00589-20>.
 25. Payros D, Secher T, Boury M, Brehin C, Menard S, Salvador-Cartier C, et al. Maternally acquired genotoxic *Escherichia coli* alters offspring's intestinal homeostasis. *Gut Microbes*. 2014;5(3):313–25. <https://doi.org/10.4161/gmic.28932>.
 26. Meester RG, Doubeni CA, Lansdorp-Vogelaar I, Goede SL, Levin TR, Quinn VP, et al. Colorectal cancer deaths attributable to nonuse of screening in the United States. *Ann Epidemiol*. 2015;25(3):208–13 e1. <https://doi.org/10.1016/j.annepidem.2014.11.011>.
 27. Park SM, Won DD, Lee BJ, Escobedo D, Esteva A, Aalipour A, et al. A mountable toilet system for personalized health monitoring via the analysis of excreta. *Nat Biomed Eng*. 2020;4(6):624–35. <https://doi.org/10.1038/s41551-020-0534-9>.
 28. Yuan X, Long Y, Ji Z, Gao J, Fu T, Yan M, et al. Green tea liquid consumption alters the human intestinal and Oral microbiome. *Mol Nutr Food Res*. 2018;62, e1800178(12). <https://doi.org/10.1002/mnfr.201800178>.
 29. Nagatake T, Kunisawa J. Emerging roles of metabolites of omega3 and omega6 essential fatty acids in the control of intestinal inflammation. *Int Immunol*. 2019;31(9):569–77. <https://doi.org/10.1093/intimm/dxy086>.
 30. Sun M, Wu W, Liu Z, Cong Y. Microbiota metabolite short chain fatty acids, GPCR, and inflammatory bowel diseases. *J Gastroenterol*. 2017;52(1):1–8. <https://doi.org/10.1007/s00535-016-1242-9>.
 31. Tilg H, Adolph TE, Gerner RR, Moschen AR. The intestinal microbiota in colorectal Cancer. *Cancer Cell*. 2018;33(6):954–64. <https://doi.org/10.1016/j.ccell.2018.03.004>.
 32. O'Keefe SJ. Diet, microorganisms and their metabolites, and colon cancer. *Nat Rev Gastroenterol Hepatol*. 2016;13(12):691–706. <https://doi.org/10.1038/nrgastro.2016.165>.
 33. Pickens CA, Lane-Elliot A, Comstock SS, Fenton JJ. Altered saturated and monounsaturated plasma phospholipid fatty acid profiles in adult males with Colon adenomas. *Cancer Epidemiol Biomark Prev*. 2016;25(3):498–506. <https://doi.org/10.1158/1055-9965.EPI-15-0696>.
 34. Hodge AM, Williamson EJ, Bassett JK, MacInnis RJ, Giles GG, English DR. Dietary and biomarker estimates of fatty acids and risk of colorectal cancer. *Int J Cancer*. 2015;137(5):1224–34. <https://doi.org/10.1002/ijc.29479>.
 35. Ohigashi S, Sudo K, Kobayashi D, Takahashi O, Takahashi T, Asahara T, et al. Changes of the intestinal microbiota, short chain fatty acids, and fecal pH in patients with colorectal cancer. *Dig Dis Sci*. 2013;58(6):1717–26. <https://doi.org/10.1007/s10620-012-2526-4>.
 36. Weir TL, Manter DK, Sheflin AM, Barnett BA, Heuberger AL, Ryan EP. Stool microbiome and metabolome differences between colorectal cancer patients and healthy adults. *PLoS One*. 2013;8(8):e70803. <https://doi.org/10.1371/journal.pone.0070803>.
 37. Clausen MR, Bonnen H, Tvede M, Mortensen PB. Colonic fermentation to short-chain fatty acids is decreased in antibiotic-associated diarrhea. *Gastroenterology*. 1991;101(6):1497–504. [https://doi.org/10.1016/0016-5085\(91\)90384-w](https://doi.org/10.1016/0016-5085(91)90384-w).
 38. Binder HJ. Role of colonic short-chain fatty acid transport in diarrhea. *Annu Rev Physiol*. 2010;72:297–313. <https://doi.org/10.1146/annurev-physiol-021909-135817>.
 39. Muller M, Hermes GDA, Canfora EE, Smidt H, Masclee AAM, Zoetendal EG, et al. Distal colonic transit is linked to gut microbiota diversity and microbial fermentation in humans with slow colonic transit. *Am J Physiol Gastrointest Liver Physiol*. 2020;318(2):G361–G9. <https://doi.org/10.1152/ajpgi.00283.2019>.
 40. Parthasarathy G, Chen J, Chen X, Chia N, O'Connor HM, Wolf PG, et al. Relationship between microbiota of the colonic mucosa vs feces and symptoms, colonic transit, and methane production in female patients with chronic constipation. *Gastroenterology*. 2016;150(2):367–79 e1. <https://doi.org/10.1053/j.gastro.2015.10.005>.
 41. Butler LM, Yuan JM, Huang JY, Su J, Wang R, Koh WP, et al. Plasma fatty acids and risk of colon and rectal cancers in the Singapore Chinese health study. *NPJ Precis Oncol*. 2017;1(1):38. <https://doi.org/10.1038/s41698-017-0040-z>.
 42. Ohno H, Murakami H, Tanisawa K, Konishi K, Miyachi M. Validity of an observational assessment tool for multifaceted evaluation of faecal condition. *Sci Rep*. 2019;9(1):3760. <https://doi.org/10.1038/s41598-019-40178-5>.
 43. Kay IT, Weimer M, Watson CJ. The formation in vitro of stercobilin from bilirubin. *J Biol Chem*. 1963;238:1122–3.
 44. Kobayashi S, Honda S, Murakami K, Sasaki S, Okubo H, Hirota N, et al. Both comprehensive and brief self-administered diet history questionnaires satisfactorily rank nutrient intakes in Japanese adults. *J Epidemiol*. 2012;22(2):151–9. <https://doi.org/10.2188/jeaj.20110075>.

45. Hirayama Y, Tsunematsu Y, Yoshikawa Y, Tamafune R, Matsuzaki N, Iwashita Y, et al. Activity-based probe for screening of high-Colibactin producers from clinical samples. *Org Lett*. 2019;21(12):4490–4. <https://doi.org/10.1021/acs.orglett.9b01345>.
46. Hosomi K, Ohno H, Murakami H, Natsume-Kitatani Y, Tanisawa K, Hirata S, 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(1):4339. <https://doi.org/10.1038/s41598-017-04511-0>.
47. Folch J, Lees M, Sloane Stanley GH. A simple method for the isolation and purification of total lipides from animal tissues. *J Biol Chem*. 1957;226(1):497–509.
48. Kundel HL, Polansky M. Measurement of observer agreement. *Radiology*. 2003;228(2):303–8. <https://doi.org/10.1148/radiol.2282011860>.
49. Venkaiah K, Brahmam GN, Vijayaraghavan K. Application of factor analysis to identify dietary patterns and use of factor scores to study their relationship with nutritional status of adult rural populations. *J Health Popul Nutr*. 2011; 29(4):327–38. <https://doi.org/10.3329/jhpn.v29i4.8448>.
50. Guadagnoli E, Velicer WF. Relation of sample size to the stability of component patterns. *Psychol Bull*. 1988;103(2):265–75. <https://doi.org/10.1037/0033-2909.103.2.265>.

Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

