

Stool pattern is associated with prevalence of tumorigenic bacteria of the microbiota and plasma and fecal fatty acids in healthy Japanese adults

Daiki Watanabe

National Institutes of Biomedical Innovation, Health and Nutrition

Haruka Murakami

National Institutes of Biomedical Innovation, Health and Nutrition

Harumi Ohno

National Institutes of Biomedical Innovation, Health and Nutrition

Kumpei Tanisawa

National Institutes of Biomedical Innovation, Health and Nutrition

Kana Konishi

National Institutes of Biomedical Innovation, Health and Nutrition

Kikue Todoroki-Mori

National Institutes of Biomedical Innovation, Health and Nutrition

Yuta Tsunematsu

University of Shizuoka

Michio Sato

University of Shizuoka

Yuji Ogata

University of Shizuoka

Noriyuki Miyoshi

University of Shizuoka

Naoto Kubota

National Institutes of Biomedical Innovation, Health and Nutrition

Jun Kunisawa

National Institutes of Biomedical Innovation, Health and Nutrition

Keiji Wakabayashi

University of Shizuoka

Tetsuya Kubota

National Institutes of Biomedical Innovation, Health and Nutrition

Kenji Watanabe

University of Shizuoka

Motohiko Miyachi (✉ miyachi@nibiohn.go.jp)

National Institutes of Biomedical Innovation, Health and Nutrition <https://orcid.org/0000-0002-1146-0905>

Keywords: tumorigenic bacteria, stool pattern, fatty acid, cross-sectional study

Posted Date: November 13th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-106339/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background: Colibactin-producing *Escherichia coli* containing polyketide synthase (*pks⁺ E. coli*) has been shown 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 by factorial analysis using a previously validated questionnaire including stool frequency, volume, color, shape, and odor. Factor scores were classified by tertile. The prevalence of *pks⁺ E. coli* was determined 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 contributed 70.1 % of all patterns seen (factor 1: lower frequency, darker color, and softer shape, factor 2: higher volume and harder 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 correlated with some plasma and fecal fatty acids. No other stool patterns were significant.

Conclusions: These results suggest that stool pattern may be useful for evaluating gut microbiota, including the presence of tumorigenic bacteria and fecal fatty acids. It may provide useful insight for effective early discovery strategies for CRC.

Background

Colorectal cancer (CRC) is the second most common cause of cancer death and the third most common cancer in the world. Among all cancers, the mortality and incidence of CRC contributes approximately 9% and 10%, respectively [1]. The majority of loss of disability-adjusted life years for 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 significant impact in 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 lead to 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 a certain gut *Escherichia coli* strain harboring the genomic island (*cjb* gene cluster) encoding polyketide synthase (*pks⁺ E. coli*) [4–11]. Colibactin has been shown to lead to genomic instability to the 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 *pks⁺ E. coli* producing colibactin in the gut microbiome may be a risk factor of CRC and may be a useful target for identifying groups at high risk for both incidence and progression.

Stool status variables, such as stool shape [12–14], frequency [12, 15], and color [16] have been associated with potentially adverse gut microbiome profiles from fecal analysis in healthy individuals and in patients with acute gastroenteritis. Using a stool pattern approach, which considers a more comprehensive overview of the stool status, could provide more interpretable findings than studying single stool exams since some stool status variables are related to each other, such as stool frequency and shape [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 participant characteristics in the analysis cohort. Of the included 224 participants, 60 participants were noted *pks⁺ E. coli* positive (26.8%). In comparing participants who were *pks⁺ E. coli* positive and negative, the presence of *pks⁺ E. coli* tended to be greater in both men and alcohol drinkers. Other variables were not significantly associated.

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

	<i>Total</i> (n = 224)		<i>pks</i> ⁺ <i>E. coli</i> (n = 60)		<i>pks</i> ⁻ <i>E. coli</i> (n = 164)		<i>p</i> -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 [μ U/mL] ^a	4.0	(3.9)	3.6	(1.8)	4.1	(4.5)	0.414

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).

^a Continuous 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)

^b Categorical variables are shown as number of individuals (%) and were analyzed using a χ^2 test.

	<i>Total</i> (n = 224)	<i>pks⁺ E. coli</i> (n = 60)	<i>pks⁻ E. coli</i> (n = 164)	<i>p</i> -value
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 <i>p</i> -values shown in bold are statistically significant (<i>p</i> < 0.05).				
^a Continuous 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 (m ²)				
^b Categorical variables are shown as number of individuals (%) and were analyzed using a χ^2 test.				

Reproducibility Of Self-reported Stool Status

We evaluated the reproducibility of the 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.

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).

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

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

Table 3
Stool patterns extracted 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	48.4	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)].			

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 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), but no significance was seen with 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).

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

Stool patterns	Number or (95% CI)			<i>p</i> for trend ^a
	Q1	Q2	Q3	
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
<i>Ref</i> reference. The prevalence rates of <i>pks</i> ⁺ <i>E. coli</i> 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 softer shape; factor 2: higher volume and harder shape; and factor 3: darker color and stronger odor.				
^a Statistical 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 <i>p</i> values are statistically significant (<i>p</i> <0.05).				
^b Model 1 was adjusted for age (continuous) and sex (male or female).				
^c Model 2 was as model 1 plus mutual adjustment for BMI (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).				

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; and was significantly negatively correlated with plasma eicosenic acid and α-linoleic acid, as well as fecal propionate and

succinate. Other stool patterns had no significant correlation. The correlations between stool status and plasma and fecal fatty acids were tabulated and presented in supplementary Tables 2 and 3, respectively.

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

	Structure	<i>n</i>	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
Eicosenic 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

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).

Structure	<i>n</i>	Unit	Median	Interquartile range	Stool status pattern		
					1	2	3
Dihomo-gammalinolenic 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	-0.08
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	0.00
Lactate	C3:0	19	mol/g	0.08 (0.03 to 0.30)	0.04	-0.08	0.44

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).

	Structure	<i>n</i>	Unit	Median	Interquartile range	Stool status pattern			
						1	2	3	
Isobutyrate	C4:0	161	mol/g	1.57	(0.95 to 2.19)	0.21	**	-0.11	-0.15
Butyrate	C4:0	161	mol/g	10.30	(6.59 to 17.52)	-0.07		0.06	-0.01
Succinate	C4:0	143	mol/g	0.18	(0.09 to 0.34)	-0.21	**	0.05	0.09
Isovalerate	C5:0	161	mol/g	1.01	(0.57 to 1.50)	0.27	**	-0.12	-0.14
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).

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 adjusting 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. As far as we know, this is the first study to show a stool pattern association with not only prevalence of *pks*⁺ *E. coli*, but also plasma and fecal fatty acids. These associations suggest that stool pattern may reflect the gut microbiota, including the presence of tumorigenic bacteria.

Certain risk factors for CRC incidence have been identified, including smoking, obesity, diabetes, and high consumption of alcohol, as well as consuming red and processed meats, in epidemiological studies [20]. These identified CRC-risk factors have not only been associated with increased CRC incidence, they have also been associated with potentially adverse gut microbiome profiles [21]. 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] compared to healthy individuals. Thus, it is important to evaluate the association between environmental exposure factors and the prevalence of tumorigenic bacteria in the gut microbiota.

Our results demonstrated a significant association between the prevalence of *pks*⁺ *E. coli* and stool pattern. Animal models mimicking the natural transmission of *E. coli* producing colibactin from mothers to neonates has shown lower rates of Firmicutes taxa, Proteobacteria taxa, and microbial species richness, as well as higher DNA repair

function compared to the sham model [24]. This model has 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* as a tumorigenic bacteria and an increased risk of CRC has not been established well, this study may underscore the 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 also suggest that stool pattern might be a marker associated with the prevalence of tumorigenic bacteria in healthy individuals. Longitudinal objective monitoring of a person's stool status from serial samples taken from an individual's excreta at home, as previously suggested [27], may be the most reasonable and cost-effective method for early detection of risk factors of CRC.

Nutrients derived from our ingested food are utilized by the gut microbiome, with certain preferred energy sources such as short chain fatty acids (SCFAs) for colonocytes [28, 29]. These metabolites can suppress inflammation and carcinogenesis through effects on immunity, gene expression, and epigenetic modulation [28–31]. Some plasma fatty acids have been noted to be inversely or positively associated with the presence of colon adenomas [32] and increased risk of CRC in middle aged adults [33]. In addition, studies in CRC patients have noted lower levels of propionate and butyrate [34] and higher levels of valeric acid, isobutyric acid, and isovaleric acid [35] in SCFAs derived from fecal samples compared to healthy controls. Production of SCFAs has been shown to be reduced in patients with diarrhea compared to those without diarrhea [36]. In addition, inhibition of SCFA synthesis by administration of polyethylglycol and antibiotics has been reported to result in diarrhea [37]. It has been reported that the distal colon transit, reflected in stool frequency, was associated with not only plasma acetate and fecal SCFAs [38], but also with microbiota diversity, especially the Firmicutes taxa (*Faecalibacterium*, *Lactococcus*, and *Roseburia*) [39]. Our results indicated that the stool pattern that showed a relationship with the prevalence of *pks⁺ E. coli* was also significantly correlated with certain plasma fatty acids, including α-linoleic acid [40] and certain fecal SCFAs, such as propionate [34] and isovaleric acid [35]. These were also associated with higher incidences of CRC, supporting previous findings. Taking information from previous studies, we speculate that gut microbiota and dietary components interacted to generate biologically active molecules including SCFAs, which influenced gut secretion and motility, and that this could play a fundamental role in stool status [30, 31], as well as affect the prevalence of *pks⁺ E. coli*. 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 could be a possible tool to assess the gut microbiota, screening for the presence of tumorigenic bacteria and specific fecal fatty acids via comprehensive examination of variables including 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 to assess the stool status had previously been validated against objective fecal characteristics as well [41]. 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 a tumorigenic bacteria and stool pattern.

However, this study has a number of methodological limitations. Even when minimizing the effect of confounders using multivariate analysis to adjust for known covariates, being a cross-sectional study, we are unable to theorize about the temporal and direct causality of the observed association between stool pattern and the prevalence of

pks⁺ *E. coli*. Second, this study detected the *c/b* gene cluster in DNA extracted from fecal samples, not from the DNA of isolated *E. coli* pure cultures. Previous study has been evaluated prevalence of *pks⁺* *E. coli* by the selective cultivation method [23]. However, our previous study [17] demonstrated that prevalence of *pks⁺* *E. coli* isolated from fecal matter are relatively similar previous reports investigating prevalence of *pks⁺* *E. coli* by the selective cultivation method [23]. Therefore, it will be necessary to evaluate the concordance rate of prevalence of *pks⁺* *E. coli* defined using these two different methods for the same subject. Third, although our results showed that a softer stool shape was negatively associated with the prevalence of *pks⁺* *E. coli*, it is unclear whether participants with diarrhea, who have softer stools, have a lower prevalence of *pks⁺* *E. coli* or whether the reverse was true. In addition, we were unable to completely exclude systematic error due to self-reporting. In addition, 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 which is the oxidized metabolite of urobilinogen [42]. Stercobilin derived from bile pigment is responsible for the brown color of human feces. Since we did not measure stercobilin and bile acids directly in all participants, we could not account for their possible effects on the results, though our results were similar after adjusting 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 NEXIS cohort study, 259 adults agreed to participate. As the participation rate was relatively low, selection bias may have occurred. In addition, participants were all living in the Tokyo metropolitan area and the mean age was 58 years old in Japan. These limitations may prevent the generalization of our results. Therefore, prospective cohort studies with larger randomized samples should be done to further investigate the association between the prevalence of *pks⁺* *E. coli* and stool pattern.

Conclusion

These results suggest that an adverse stool pattern is positively associated with prevalence of *pks⁺* *E. coli*. Given the rapidly increasing incident and mortality rates for CRC around the world, its early discovery is important for both enabling people to stay healthy and for limiting the burden of healthcare-related costs. Therefore, stool pattern may be useful for evaluating gut microbiota including the presence of tumorigenic bacteria and fecal fatty acids, and it may provide useful insight for effective early discovery strategies for CRC.

Methods

Participants and study procedure

This cross-sectional study utilized data from the Nutrition and Exercise Intervention Study (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, 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 NIBIOHN. After study procedures and the risks associated with participation in this study were explained, written informed consent

was obtained from all participants before data acquisition. This study was carried out 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 [43]. 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. There, they received physical and health examinations such as anthropometry and blood tests. Investigators, registered dieticians, 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, 44]. To confirm the presence of *pks⁺* *E. coli*, we performed PCR to amplify the genes from the *cjb* cluster using bacterial genomic DNA as a template. We used the PrimeSTAR GXL DNA polymerase or the SapphireAmp Fast PCR Master Mix (Takara Bio Inc., Shiga, Japan) for PCR analysis. The primers used in the PCR experiments are as follows: *cjbB* forward primer: 5'-tgtccgtttgtggttcagcg-3', reverse primer: 5'-gtgcgctgaccattgaagattccg-3'; *cjbJ* forward primer: 5'-tggcctgtattgaaagagaccacggt-3', reverse primer: 5'-aatggAACGGTTGATGACGTCT-3'; *cjbQ* forward primer: 5'-ctgtgtcttacgatggatgccg-3', reverse primer: 5'-gcattaccagattgtcagcatgcc-3'. In this analysis, samples amplified with the appropriate amplicon length in the three *cjb* genes were defined as *pks⁺* *E. coli* positive individuals.

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 [41]. 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 plasma following the methodology reported by Folch et al. [45]. After hydrolysis with KOH, fatty acids were extracted with hexane and tricosanoic acid (C23:0) as an internal standard. Methyl esterified fatty acids were prepared with trimethylsilylating reagent and subjected to gas chromatography (GC). GC-electrospray ionization mass spectrometry (ESI/MS) analysis was performed using a Hitachi 063 gas chromatograph (Hitachi High-Tech Corporation, Tokyo, Japan) equipped with a hydrogen flame ionization detector. A glass column, 40 m × 0.3 mm in volume, was coated with diethylene glycol succinate (DEGS). Nitrogen gas was used as a carrier gas and delivered at a flow rate of 25 mL min⁻¹. The column, detector, and injection port temperature were 180 °C, 220 °C, and 260 °C, respectively. This analysis measured the 24 types of plasma fatty acids with chain lengths from 12 to 24 carbons.

Measurement Of Fecal Short-chain Fatty Acids

Five to 10 mg feces were mixed with 90 uL MilliQ and 10 uL 2 mM internal standard containing acetic acid, butyric acid, and crotonic acid for 5 min. The mixture was homogenized with 50 uL HCL and 200 uL diethyl ether, and centrifuged at 3,000 rpm for 10 min at room temperature. Eighty uL of the supernatant organic layer was transferred to a new glass vial and combined with 16 uL N-tert-butyldimethylsilyl-N-methyltrifluoroacetamide (MTBSTFA) 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 hours 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 auto injector (Shimadzu Corporation, Kyoto, Japan). The capillary column was a BPX5 column (0.25 mm × 30 m × 0.25 um; 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 30:1). The injection port and interface temperature were 230 °C and 260 °C, respectively. This analysis measured the 10 types of fecal SCFAs (C1:0-C6:0).

Statistical analysis

Participant characteristics were compared between participants with and without *pks⁺ E. coli*. The variables compared were decided in accordance to baseline characteristics used in a previous study [17]. Continuous variables were shown as mean and standard deviation, with differences between the two groups evaluated using the unpaired t-test. Categorical variables were shown in 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 number and percentage. 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 [46].

To extract the primary stool patterns, we used factorial analysis with varimax rotation (orthogonal transformation) to derive non-correlated factors [47]. 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 [47]. For these reasons, we identified three stool patterns of interest from the five stool statuses. We considered stable factor load to be scores greater than 0.4 [48]. For every participant,

we calculated factor scores for each of the three retained factors by summing scores multiplied by factor load across all stool statuses. Participants were classified into tertiles (Qs) of factor score for each stool pattern.

The prevalence of *pks⁺ E. coli* in each tertile for stool pattern was shown as number of cases and percentage. To adjust the 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 BMI (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 between plasma and fecal fatty acids 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).

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.

Consent for publication

Not applicable

Availability of data and materials

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

Competing interests

The authors declare no competing interests.

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.

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, NM, KW, and KW contributed to the analysis of the evaluation of the *pks⁺*

E. coli carrier. KT, 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.

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.

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.
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.
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.
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.
5. Li ZR, Li J, Cai W, Lai JYH, McKinnie SMK, Zhang WP, et al. Macroyclic colibactin induces DNA double-strand breaks via copper-mediated oxidative cleavage. *Nat Chem*. 2019;11(10):880-9.
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.
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.
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.
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.
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).
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).
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.
13. Vandepitte 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.
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.

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.
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).
17. Watanabe D, Murakami H, Ohno H, Tanisawa K, Konishi K, Tsunematsu Y, 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(1):15221.
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.
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.
20. Brenner H, Kloost M, Pox CP. Colorectal cancer. *Lancet.* 2014;383(9927):1490-502.
21. Brennan CA, Garrett WS. Gut Microbiota, Inflammation, and Colorectal Cancer. *Annu Rev Microbiol.* 2016;70:395-411.
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.
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.
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).
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.
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.
27. Park SM, Won DD, Lee BJ, Escobedo D, Esteve 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.
28. Tilg H, Adolph TE, Gerner RR, Moschen AR. The Intestinal Microbiota in Colorectal Cancer. *Cancer Cell.* 2018;33(6):954-64.
29. O'Keefe SJ. Diet, microorganisms and their metabolites, and colon cancer. *Nat Rev Gastroenterol Hepatol.* 2016;13(12):691-706.
30. 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.
31. 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.
32. Pickens CA, Lane-Elliott A, Comstock SS, Fenton JL. Altered Saturated and Monounsaturated Plasma Phospholipid Fatty Acid Profiles in Adult Males with Colon Adenomas. *Cancer Epidemiol Biomarkers Prev.* 2016;25(3):498-506.
33. 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.

34. 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.
35. 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.
36. 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.
37. Binder HJ. Role of colonic short-chain fatty acid transport in diarrhea. *Annu Rev Physiol.* 2010;72:297-313.
38. Muller M, Hermes GDA, Canfora EE, Smidt H, Mascllee 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.
39. 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.
40. 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.
41. 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.
42. Kay IT, Weimer M, Watson CJ. The formation in vitro of stercobilin from bilirubin. *J Biol Chem.* 1963;238:1122-3.
43. 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.
44. 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.
45. 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.
46. Kundel HL, Polansky M. Measurement of observer agreement. *Radiology.* 2003;228(2):303-8.
47. 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.
48. Guadagnoli E, Velicer WF. Relation of sample size to the stability of component patterns. *Psychol Bull.* 1988;103(2):265-75.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementarymaterialcolibactinandstoolpattern.docx](#)
- [Supplementarymaterialcolibactinandstoolpattern.docx](#)