



Interplay between colorectal cancer-related lifestyles and the gut microbiome: an exploratory analysis of metagenomic data

Rilla Tammi^{1,2,18} · Mirkka Maukonen^{1,3} · Niina E. Kaartinen¹ · Kari Koponen^{1,4,5} · Teemu Niiranen^{1,6,7} · Guillaume Méric^{8,9,10,11} · Demetrius Albanes¹² · Johan G. Eriksson^{3,13} · Pekka Jousilahti^{1,2} · Seppo Koskinen¹ · Anne-Maria Pajari¹⁴ · Rob Knight¹⁵ · Aki S. Havulinna^{16,17} · Veikko Salomaa^{1,7} · Satu Männistö¹

Received: 30 September 2025 / Accepted: 23 February 2026 / Published online: 11 March 2026
© The Author(s) 2026

Abstract

Purpose The gut microbiome may modify the associations between lifestyle factors and colorectal cancer (CRC) risk, but their complex interplay, including the interactions between lifestyle factors, remain underexplored. We examined associations between CRC-related lifestyle patterns and gut microbiome diversity and composition in Finnish adults.

Methods Our data included 1,228 adults aged 25–64 years from the National FINRISK/FINDIET 2002 Study. Information on lifestyle and background factors was obtained through self-administered questionnaires. Dietary data were gathered using a 48-h dietary recall. CRC-related lifestyles were modelled using a CRC lifestyle index based on nine major risk factors for CRC. Lower index points reflected higher-risk lifestyles. The gut microbiome profiles were analyzed using shallow shotgun metagenome sequencing. Associations between the index and microbial diversity and composition were assessed using, e.g., linear regression and permutational multivariate ANOVA adjusted for relevant confounders.

Results The index explained 0.2% of the variation in microbial composition between participants ($p < 0.05$). Higher-risk lifestyles for CRC were associated with lower microbial diversity (β 0.037, p 0.009). Higher-risk lifestyles were also associated with a higher relative abundance of species representing primarily the family Lachnospiraceae and genera such as *Dorea* and *Mediterraneibacter*, and lower relative abundance of species within the genus *Bifidobacterium* (< 0.0001).

Conclusions Participants with higher- and lower-risk lifestyles showed clear differences in their gut microbiome diversity and composition, higher-risk lifestyles being associated with potentially adverse microbial traits. These findings contribute to identifying microbial features that may characterize early stages of CRC development in individuals with high-risk lifestyles.

Keywords Diet · Epidemiology · Lifestyle index · Obesity · Physical activity · Prevention

Introduction

Colorectal cancer (CRC) is one of the most common cancer types and leading causes of cancer mortality worldwide [1]. Its global incidence has increased substantially over recent decades and continues to rise, particularly in younger age groups.

Epidemiological findings indicate that lifestyle and anthropometric factors contribute significantly to CRC development; however, the biological mechanisms underlying these associations remain unclear [2, 3]. At the same time, disturbances in the gut microbiome composition and function have been linked to CRC pathogenesis, suggesting a potential pathway between the risk factors and disease onset

[4, 5]. Indeed, several major risk factors for CRC, such as obesity, physical inactivity, and high consumption of red and processed meat, have been associated with potentially carcinogenic changes in the gut microbiome [5].

While prior research has identified associations between individual CRC risk factors and the gut microbiome, combinations of risk factors, and their potential interactions, have been less explored. Observational studies have suggested that combinations of certain risk factors increase CRC risk more than individual factors alone, with some evidence of synergy [6–9]. For example, in a study of 170,000 Finnish adults, pairwise combinations of major risk factors, such as alcohol consumption and overweight, were associated with a higher CRC risk compared with individual factors [6]. Alcohol and smoking, in particular, were jointly associated with a higher

Extended author information available on the last page of the article

CRC risk in women than anticipated based on their individual associations. As the risk factors tend to cluster and may interact in complex ways, their combined associations with the gut microbiome need to be explored. Furthermore, in population-based studies, only a small proportion of microbiome variation has been attributed to individual lifestyle factors such as diet, whereas the influence of external factors combined appears to surpass that of genetic factors [10–12]. Consequently, considering combinations of risk factors may provide a more comprehensive perspective on the interplay between lifestyles, the gut microbiome, and CRC risk.

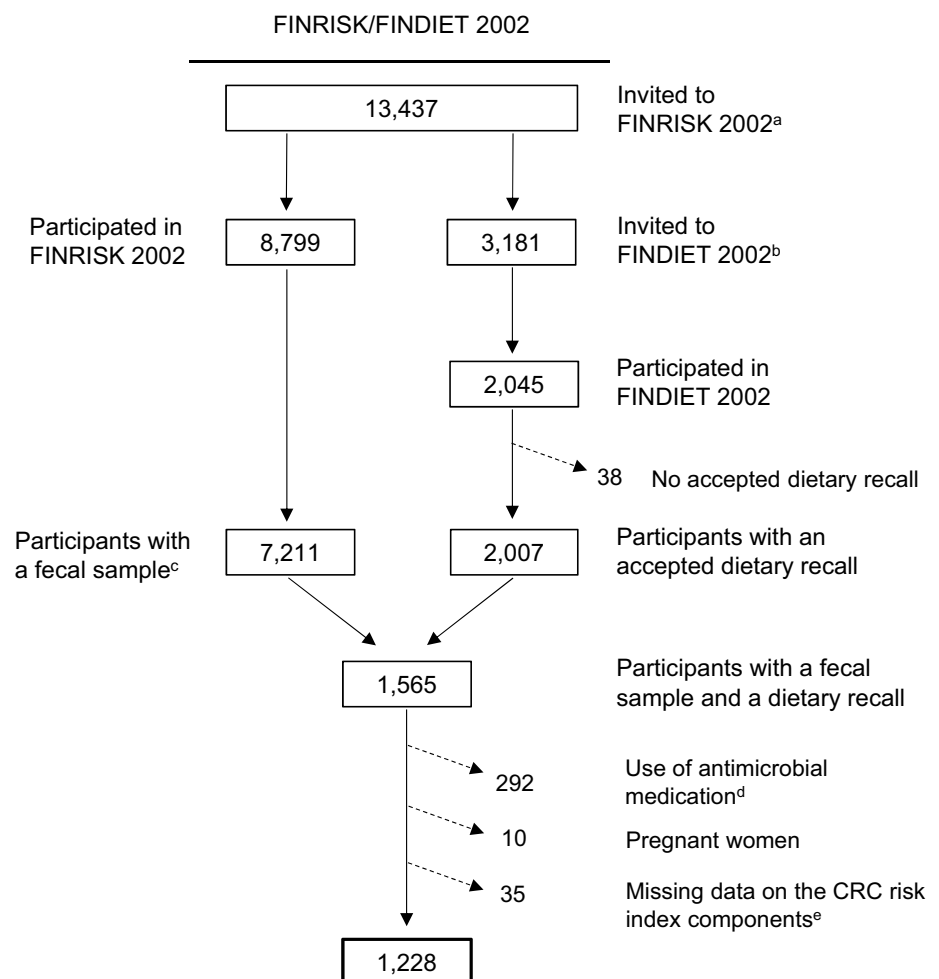
In this study, we aimed to account for the complex interplay between lifestyle and anthropometric factors and the gut microbiome in the context of CRC risk. To achieve this, we studied the associations between CRC-related lifestyles and gut microbiome diversity and composition.

Methods

Participants

Our data comprised participants from the National FINRISK/FINDIET 2002 Study (FINRISK/FINDIET 2002) [13]. FINRISK 2002 is part of the FINRISK study series conducted every 5 years from 1972 to 2012 by the Finnish Institute for Health and Welfare (THL) to monitor health behavior and risk factors of chronic diseases in Finnish adults. In these studies, data were collected in a health examination and through self-administered questionnaires. In FINRISK 2002, participants were also asked to donate a fecal sample. Additionally, FINRISK 2002 included a sub-study called FINDIET 2002, in which diet was assessed using a 48-hour dietary recall (48-h recall) [14]. In the current study, we included participants with a fecal sample and an accepted dietary recall (Fig. 1). We excluded participants with a registered purchase of systemic antibacterial medication [Anatomical Therapeutic Chemical (ATC) Classification Code J01] within the 6 months preceding the baseline

Fig. 1 Exclusions and the final study sample in FINRISK/FINDIET 2002. ^aA random sample of adults aged 25–74 years from six large geographical areas of mainland Finland. ^bThe FINDIET subsample was randomly selected from individuals invited to FINRISK 2002. ^cSuccessfully sequenced fecal sample with sufficient read count ($\geq 50,000$). ^dRegistered purchase of systemic antibacterial medication (Anatomical Therapeutic Chemical classification code J01) within 6 months prior to the baseline examination. ^eMissing data on waist circumference $n=1$, physical activity $n=6$, and alcohol intake $n=28$



examination, pregnant women, and those with missing information on the risk factors included in the CRC lifestyle index. The final study population included 1,228 participants aged 25–74 years.

Additionally, we used pooled data from five large cohorts of Finnish adults to test the performance of the CRC lifestyle index in relation to CRC risk, prior to other analyses. These cohorts included the Alpha-Tocopherol Beta-Carotene Cancer Prevention Study (ATBC) [15], the Health 2000 Health Examination Survey (Health 2000) [16], the Helsinki Birth Cohort Study (HBCS) [17], the DIetary, Lifestyle and Genetic Determinants of Obesity and Metabolic Syndrome 2007 Study (DILGOM 2007) [18], and the National FINRISK 2012 Study (FINRISK 2012) [13] (Online Resource 1). In each cohort, data were collected through a health examination and self-administered questionnaires, including a food frequency questionnaire (FFQ). The cohorts were followed using national health registers. The final study sample included 43,488 participants aged 25–99 years (Online Resource 2). The methods described below apply to FINRISK/FINDIET 2002 and the pooled cohorts unless otherwise indicated.

CRC lifestyle index

To model CRC-related lifestyles, we adapted the standardized 2018 World Cancer Research Fund/American Institute for Cancer Research (WCRF/AICR) Score [19] to include only CRC-specific risk factors. The standardized score operationalizes the general cancer prevention recommendations of the 2018 WCRF/AICR Third Expert Report [20]. Based on the same report, we included in the index nine lifestyle and anthropometric factors with strong evidence of causality for CRC. These were body mass index (BMI), waist circumference (WC), height, leisure-time physical activity, consumption of whole grains, dairy products, and red and processed meat, and intake of alcohol (Table 1). Each component was assigned 0, 0.5, or 1 point representing not meeting, partially meeting, or meeting the target cut-off [19]. The cut-offs were determined based on the literature, primarily the 2018 WCRF/AICR Report (Table 1) [20]. As exceptions, height was scored using sex-specific tertiles because cut-offs could not be determined, and physical activity was scored based on predefined categories of the variable available in our data.

As per the standardized score, we averaged the points for BMI and WC to form one body-fatness score. In one of the pooled cohorts (ATBC), WC was not measured, for which the score was based on BMI alone [19]. To avoid overemphasizing diet, we also averaged the points for the four dietary components to form a single dietary score. Finally, the points for body fatness, height, physical activity, and diet were summed, resulting in scores ranging between zero and

four points, with higher points indicating fewer CRC risk factors (i.e., lower-risk lifestyles for CRC).

From the strong-evidence risk factors identified in the 2018 WCRF/AICR Report [20], we omitted dietary fiber to avoid redundancy with whole grains. Whole grains were prioritized because the evidence of associations with a reduced CRC risk has been strongest for cereal fiber [21, 22], and whole grains are a major source of both cereal fiber and total fiber in Finland [23]. Moreover, the protective effects of whole grains may extend beyond the role of fiber per se, owing to their various potentially beneficial bioactive compounds [24, 25]. We also omitted the use of calcium supplements because of the high dairy product consumption in Finland, for which dietary calcium intake is generally sufficient in adults [23, 26].

Dietary factors

In FINRISK/FINDIET 2002, detailed dietary information was collected using a 48-h recall capturing food consumption over the two preceding days [26]. Trained nutritionists interviewed participants during the health examination, documenting all foods and beverages consumed within the study period. Portion sizes were estimated based on a validated portion size picture booklet [31], household measures, and standard food packaging. To estimate alcohol intake, we used data collected through self-administered questionnaires to better capture habitual use. The data included the number of portions and frequency of consumption of different alcoholic beverages over the past year.

In the pooled cohorts, dietary information was collected through validated semi-quantitative FFQs that assessed habitual food consumption over the past year [32–35]. The FFQ included approximately 130 food items and mixed dishes (276 in ATBC), and the consumption of each item was recorded by up to nine frequency categories and fixed portion sizes.

The average daily food consumption (g/day) and energy intake (kJ/day) were calculated using the Finnish Food Composition Database Fineli® and an in-house software [36]. Weekly consumption was calculated by multiplying daily consumption by seven. Food consumption was calculated at the ingredient level by breaking down foods and mixed dishes into basic ingredients using Fineli® standard recipes. Whole grain consumption was calculated as the sum of rye, oat, and barley consumption, which has been shown to correlate strongly with total whole grain intake in Finnish adults ($r=0.99$) [29]. Dairy products included milk, fermented dairy products (e.g., yoghurt), cheese, cream, and ice cream. The contribution of ice cream on dairy consumption was small. Red meat included beef,

Table 1 The components and scoring of the CRC lifestyle index

Component	Scoring		
	0	0.5	1
Body fatness ^a			
BMI ^b (kg/m ²)	≥ 30	25–< 30	< 25
WC ^b (cm)			
Women	≥ 88	80–< 88	< 80
Men	≥ 102	94–< 102	< 94
Height ^c (cm)	Tertile 3	Tertile 2	Tertile 1
Leisure-time physical activity ^d	Inactive	Moderately active	Active
Diet ^e			
Whole grains ^{f,g} (g/day)	< 45	45–< 90	≥ 90
Dairy products ^g (g/day)	< 200	200–< 400	≥ 400
Red and processed meat ^h (g/week)	Red meat > 350 or processed meat ≥ 100	Red meat ≤ 350 and processed meat 21–< 100	Red meat ≤ 350 and processed meat ≤ 21
Alcohol (100%) ⁱ (g/day)	> 20	> 0–≤ 20	0

BMI body mass index, *WC* waist circumference

^aThe points for BMI and WC were averaged to calculate one score for body fatness. Body fatness was based on BMI alone if data on WC was missing in the entire dataset

^bThe cut-offs are based on the 2018 WCRF/AICR Third Expert Report [20] and the WHO guidelines [27, 28]

^cHeight was scored based on sex-specific tertiles. Tertile medians (cm) in FINRISK/FINDIET 2002 were in women T1: 157, T2: 163, T3: 169; in men T1: 170, T2: 176, T3: 183

^dLeisure-time physical activity was scored based on predefined categories of the variable available in our data: inactive (only light activities, e.g., reading), moderately active (e.g., walking ≥ 4 h/week), or active (e.g., running or competitive sports ≥ 3 h/week)

^eThe points for the consumption of whole grains, dairy products, and red and processed meat, and intake of alcohol were averaged to calculate one dietary score

^fWhole grain consumption was calculated as the sum of rye, oat and barley consumption [29]

^gThe cut-off for meeting the target consumption was based on the 2018 WCRF/AICR Third Expert Report [20]. The cut-off for partially meeting the target consumption was defined as consuming at least half of the target amount [19]

^hThe cut-offs are based on the 2018 WCRF/AICR Third Expert Report [20], the standardized 2018 WCRF/AICR Score [19] and the Nordic Nutrition Recommendations 2023 [30]

ⁱThe cut-offs are based on the 2018 WCRF/AICR Third Expert Report [20]

pork, and lamb. Processed meat included sausages and cold cuts.

Physical activity and anthropometric factors

Information on leisure-time physical activity was collected through self-administered questionnaires. Based on the collected data, participants were categorized into inactive (only light activities, e.g., reading), moderately active (e.g., walking ≥ 4 h/week), and active (e.g., running and competitive sports ≥ 3 h/week) [37].

Weight (kg), height (m), and WC (cm), were measured in the health examination by trained research staff using

international standard protocols [38]. BMI was calculated as weight divided by height squared (kg/m²).

Confounding factors

Information on participants' sex and age originated from the sampling frame (Finnish Population Information System). Information on educational attainment, smoking history, and use of hormone-replacement therapy (HRT) were collected through self-administered questionnaires. Educational attainment was categorized into low, middle, or high. Smoking status was categorized into current, former, or never smoker. In ATBC, all participants were male and current smokers at baseline owing to the study design. In FINRISK/FINDIET 2002, current smokers were combined with former

smokers who had smoked within the past 6 months ('current smokers') and never smokers with those who had not smoked within the past 6 months ('nonsmokers') to create a binary variable. HRT use (in women) was categorized as 'ever' or 'never'. Educational attainment and HRT use were adjusted for only in testing the CRC index.

Information on the use of potentially microbiome-altering medication was obtained from the prescription medicine purchase register maintained by the Social Insurance Institution of Finland. In addition to systemic antibacterial medication, we considered metformin (ATC code A10BA02), psycholeptics (NO5), psychoanaleptics (NO6), proton pump inhibitors (A02BC), and constipation medication (A06A).

Information on baseline diabetes status was obtained from the nationwide administrative registries on reimbursement for diabetes medication expenses and medication purchases (ATC codes A10), hospitalizations, or causes of death (International Classification of Diseases 10 (ICD-10) codes E10–14) [39]. Participants were linked to the registries using unique personal identity codes issued to all Finnish citizens and permanent residents. Diabetes status was adjusted for in testing the CRC index, but not in the microbiome analyses due to overlap with metformin use.

Colorectal cancer ascertainment

Data on CRC diagnoses (ICD-9 codes 153, 154.0, 154.1 or ICD-10 codes C18, C19.9 and C20.9) were obtained from the Finnish Cancer Registry [40]. Participants were linked to the registry using the unique personal identity codes. The follow-up extended from the date of enrolment until CRC diagnosis, death, or the end of study period (see Online Resource 1 for the pooled cohorts).

The final FINRISK/FINDIET 2002 study sample did not include participants with prevalent CRC at baseline. During the follow-up (until 31 December 2022), 12 participants were diagnosed with CRC. Of these, all but one were diagnosed more than 8 years after the baseline.

Microbiome characterization

All participants in FINRISK/FINDIET 2002 willing to donate a fecal sample received a sampling kit during the health examination with detailed instruction on its use. The participants were instructed to gather the sample at home at their earliest opportunity and mail it overnight to the THL Laboratory. The samples were collected into 50-ml Falcon tubes without a stabilizing solution and mailed under typical Finnish winter conditions. Upon receipt, the samples were frozen and stored at -20°C until sequencing in 2017 at the University of California San Diego.

A detailed description of the DNA extraction and library preparation has been published elsewhere [41]. The samples

were characterized using shallow shotgun whole-metagenome sequencing with the Illumina Hi-Seq 4000 instrument (Illumina, Inc.), generating 150-bp paired-end reads [42]. The mean read count per sample was approximately 900,000. Adapter sequences were removed using fastp (version 0.23.4) [43]. Host sequences were filtered by mapping reads against the human pangenome, T2T-CHM13v2.0, and GRCh38 using minimap2, followed by extraction of unmapped reads with samtools [44]. Raw sequences were then taxonomically annotated using incorporated metagenomes in Greengenes2 [45].

We measured alpha diversity, which quantifies within-individual microbial diversity, using the Shannon index based on species-level raw counts [46]. Beta diversity, which quantifies between-individual variation in microbial community composition, was measured using the weighted UniFrac metric based on relative abundances [47]. Relative abundances were calculated from the annotated data by scaling the number of raw read counts of each taxon to the total sum of reads. Differential abundance analyses were conducted using species-level raw counts, excluding rare species with a prevalence $< 1\%$ or a relative abundance $< 0.01\%$ [41].

Statistical analyses

Throughout the study, continuous variables were characterized using medians and interquartile ranges (IQR) and categorical variables using proportions (%). We tested the CRC lifestyle index in relation to CRC risk in the pooled cohorts using two-stage meta-analysis. The index was divided into quintiles cohort-specifically. First, we calculated cohort-specific hazard ratios (HR) and 95% confidence intervals (95% CI) using multivariate Cox regression models [48]. The estimates were subsequently pooled using random-effects models, with weights equal to the inverse of their variances. The proportional hazards assumption was met. p for trend over the index quintiles was calculated using the quintile medians as continuous independent variables. Heterogeneity between the cohorts was tested using Q -statistics. The cohort-specific analyses were adjusted for sex, age, energy intake, educational attainment, and smoking status. Additionally, three sensitivity analyses were conducted: (1) excluding participants diagnosed with CRC within the first 2 years of follow-up, (2) further adjusting for prevalent diabetes, and (3) further adjusting for HRT use in women. We examined associations between the index components and CRC risk using the same methods.

In the microbiome analyses, we used multiple linear regression to examine associations of the continuous CRC lifestyle index and its components with alpha diversity. To analyze associations between the index and beta diversity, we used Permutational Multivariate Analysis of Variance (PERMANOVA) [49], dispersion analysis, Distance-based

Redundancy Analysis (dbRDA) [50], and Principal Coordinates Analysis (PCoA). PERMANOVA was used to test the degree to which the CRC index (and its components) explain the compositional variation in microbiome between individuals [49]. Dispersion analysis was used to test whether differences in beta diversity across the index were driven by within-group variability. DbRDA was used to examine constrained variance, i.e., how much the index and confounders together explain of the variation in beta diversity, and visualize the direction of associations between them. PCoA was used to visualize the clustering of microbial communities in the highest and lowest index quintile, and the function ‘factorfit’ from the package *vegan* to assess the significance of the clustering. PERMANOVA, dbRDA and factorfit were run with 999 permutations.

To study associations between the CRC lifestyle index and all taxa at species level, we used Analysis of Compositions of Microbiomes with Bias Correction 2 (ANCOM-BC2) [51]. As a secondary analysis to further examine species co-occurrence, we clustered the species that were significantly associated with the index and passed the ANCOM-BC2 robustness screening, using Ward’s minimum variance method based on proportionality. The optimal number of clusters was determined using Kelley–Gardner–Sutcliffe penalty function [52]. The results were visualized with a heatmap.

The microbiome analyses were adjusted for sex, age, energy intake, smoking status, use of potentially microbiome-altering medication, and BMI (with predictors other than the index, BMI, or WC). Interaction between the index and sex was tested by including an interaction term in the analyses. In the linear regression analyses (alpha diversity) and PERMANOVA (beta diversity), *p* values were corrected

for multiple testing using the Benjamini–Hochberg false discovery rate (FDR) procedure [53]. In the differential abundance analyses, multiple testing was controlled internally by ANCOM-BC2 [51]. A two-sided (FDR-corrected) *p* value of <0.05 was used to indicate statistical significance.

All analyses were conducted and figures generated using R statistical software versions 4.3 and 4.4 [54]. Key packages included *meta* (8.0.1), *mia* (1.12.0), *vegan* (2.6.8), and *microbiome* (1.28.0) [55–58].

Results

Testing the CRC lifestyle index

In the pooled cohorts, 79% of the participants were men, and the median follow-up was 14 years (range 7.8–17.9 years), during which 1,118 CRC cases were diagnosed (Online Resource 1). The median age ranged from 50 to 60 years (Online Resource 3). The median points for the CRC lifestyle index ranged from 1.75 to 2.00.

Participants with the highest points in the CRC lifestyle index (Q5) had a 31% lower CRC risk compared with those with the lowest points (Q1, HR 0.69, 95% CI 0.58–0.83; Table 2). Each one-point increase in the index was associated with a 19% decrease in CRC risk (0.81, 0.74–0.88). A one-point increase corresponds to, for example, the difference between being physically active vs. inactive or having both BMI and WC in the lowest vs. highest category. There was no statistically significant heterogeneity between the cohorts. Furthermore, we observed no significant interaction by sex, and the results were similar in men and women. The results remained similar also in the sensitivity analyses

Table 2 Associations (HR and 95% CI) between the CRC lifestyle index (in quintiles and continuously) and CRC risk in the pooled cohorts

	Index quintiles ^a			<i>p</i> _{trend} ^b	<i>p</i> _{het.} ^c	Continuous index ^d	<i>p</i>
	Q1	Q3	Q5				
Number of participants	8,271	8,579	9,511			43,488	
Number of CRC cases	251	196	234			1,118	
HR (95% CI) ^e	1.00	0.71 (0.59–0.85)	0.69 (0.58–0.83)	<0.0001	0.44	0.81 (0.74–0.88)	<0.0001

The pooled cohorts included the Alpha-Tocopherol Beta-Carotene Cancer Prevention Study (ATBC), the Health 2000 Health Examination Survey (Health 2000), the Helsinki Birth Cohort Study (HBCS), the Dietary, Lifestyle and Genetic Determinants of Obesity and Metabolic Syndrome 2007 Study (DILGOM 2007), and the National FINRISK 2012 Study (FINRISK 2012)

CRC colorectal cancer, HR hazard ratio, 95% CI 95% confidence intervals

^aMedian index points in the CRC lifestyle index quintiles in ATBC Q1: 1.000, Q3: 1.875, Q5: 2.625, in Health 2000 Q1: 1.000, Q3: 1.875, Q5: 2.750, in HBCS Q1: 0.875, Q3: 1.750, Q5: 2.750, in DILGOM 2007 Q1: 1.125, Q3: 2.125, Q5: 3.000, in FINRISK 2012 Q1: 1.125, Q3: 2.000, Q5: 3.000

^b*p* for trend was calculated using index quintile medians as continuous independent variables

^c*p* for heterogeneity between the pooled cohorts was calculated by *Q*-statistics

^dHR (95% CI) per one-point increase in the index

^eAdjusted for sex, age (continuous), energy intake (continuous), educational attainment (low/middle/high), and smoking status (never/former/current smoker)

(data not shown). Of the individual index components, BMI (1.13, 1.05–1.23; per 5 kg/m²), red meat consumption (1.17, 1.00–1.38; per 100 g/day), and alcohol intake (1.03, 1.01–1.04; per 5 g/day) were significantly associated with CRC risk. Overall, these results suggest that the CRC lifestyle index is an appropriate method to distinguish between participants with higher- and lower-risk lifestyles for CRC in Finnish adults.

The CRC lifestyle index and gut microbiome diversity and composition

In FINRISK/FINDIET 2002, 46% of the participants were men and the median age was 48 years (Table 3). The median score for the CRC lifestyle index was 2.0 points, ranging from 1.1 in Q1 to 3.0 in Q5. No participant scored full points. Participants with higher index points tended to have a higher energy intake and were less often current smokers or users of potentially microbiome-altering medication than those with lower points.

The CRC lifestyle index was positively associated with alpha diversity (β 0.037, SE 0.013, p 0.009; Table 4). Of the index components, a higher BMI and WC were associated with a lower alpha diversity (BMI: β -0.045, SE 0.013, p 0.002; WC: β -0.048, SE 0.015, p 0.003), whereas being physically active versus inactive was associated with a higher diversity (β 0.091, SE 0.039, p 0.042). We observed no significant associations between the other components and alpha diversity.

The CRC lifestyle index explained a small but statistically significant proportion of the variation in beta diversity (R^2 0.002, p 0.010; Table 4). Of the index components, all except for physical activity and the consumption of red meat and dairy products, were significantly associated with beta diversity. The explained proportions ranged from 0.1 to 0.6%, the largest proportions being for BMI, WC, and alcohol intake (0.5–0.6%). In dbRDA, the index together with the confounding variables (constrained variance) explained 1.4% (p 0.001) of the between-sample variation in beta diversity. The first two axes accounted

Table 3 Participant characteristics in FINRISK/FINDIET 2002 (medians [IQR] or proportions [%])

	Total <i>n</i> = 1,228	CRC lifestyle index quintiles				
		Q1 <i>n</i> = 252	Q2 <i>n</i> = 285	Q3 <i>n</i> = 248	Q4 <i>n</i> = 234	Q5 <i>n</i> = 209
Background factors						
Men (%)	46	48	42	42	50	47
Age (years)	48 (18)	49 (17)	47 (17)	47 (18)	49 (19)	47 (21)
Energy intake (MJ/day)	7.6 (3.2)	7.3 (2.8)	7.3 (3.1)	7.8 (3.6)	7.6 (3.3)	7.8 (3.1)
Current smoker (%)	24	26	30	27	19	17
Medication use ^a (%)	8.6	11.5	9.5	11.3	5.6	4.3
CRC lifestyle index and its components						
CRC lifestyle index ^b (points)	2.0 (0.9)	1.1 (0.4)	1.8 (0.3)	2.1 (0.3)	2.5 (0.3)	3.0 (0.4)
Body fatness score (points)	0.50	0	0.50	0.75	0.75	1.00
BMI (kg/m ²)	26 (6)	30 (6)	27 (6)	26 (5)	25 (5)	24 (3)
WC (cm)						
Women	82 (16)	96 (16)	85 (13)	80 (14)	77 (10)	75 (8)
Men	95 (15)	107 (15)	97 (11)	93 (14)	92 (11)	87 (9)
Height (cm)						
Women	163 (8)	167 (5)	163 (7)	162 (9)	162 (7)	159 (5)
Men	176 (9)	181 (6)	178 (7)	177 (10)	173 (9)	172 (5)
Inactive in leisure time (%)	19	41	24	14	9	0
Dietary score (points)	0.375	0.25	0.375	0.50	0.50	0.50
Whole grains (g/day)	49 (58)	35 (48)	44 (54)	52 (56)	59 (68)	67 (65)
Dairy products (g/day)	375 (371)	307 (307)	361 (374)	387 (342)	440 (360)	416 (405)
Red meat (g/week)	232 (500)	282 (605)	264 (447)	242 (413)	243 (500)	146 (372)
Processed meat (g/week)	175 (403)	196 (414)	189 (396)	140 (421)	174 (414)	169 (420)
Alcohol (100%, g/day)	6.3 (15)	9 (22)	6 (14)	6 (13)	6 (14)	5 (10)

BMI body mass index, CRC colorectal cancer, IQR interquartile range, WC waist circumference

^aUse of metformin, psycholeptics, psychoanaleptics, proton pump inhibitors or constipation medication

^bHigher points indicate a lower-risk lifestyle for CRC

Table 4 Associations of the CRC lifestyle index and its components with alpha diversity measured using the Shannon Index and beta diversity measured using the weighted UniFrac metric

	Alpha diversity Shannon Index			Beta diversity weighted UniFrac	
	β	SE	p^a	R^2	p^a
CRC risk index ^{b,c}	0.037	0.013	0.009	0.002	0.010
Index components ^d					
Body fatness score	0.054	0.013	<0.001	0.004	<0.001
BMI (kg/m ²)	−0.045	0.013	0.002	0.005	<0.001
WC (cm)	−0.048	0.015	0.003	0.006	<0.001
Height (cm)	0.023	0.019	0.33	0.002	0.039
Leisure-time physical activity (%)					
Inactive vs. active	0.091	0.039	0.042	0.003	0.10
Dietary score	0.008	0.013	0.69	0.002	0.007
Whole grains (g/day)	−0.021	0.013	0.21	0.002	0.010
Dairy products (g/day)	0.008	0.013	0.75	0.001	0.052
Red meat (g/week)	−0.015	0.013	0.37	0.0007	0.62
Processed meat (g/week)	−0.003	0.013	0.86	0.002	0.003
Alcohol (100%) (g/day)	−0.018	0.014	0.31	0.005	<0.001

BMI body mass index, CRC colorectal cancer, SE standard error, WC waist circumference

^a p values were corrected for multiple testing using the Benjamini–Hochberg false discovery rate (FDR)[53]

^bHigher points indicate a lower-risk lifestyle for CRC

^cAdjusted for sex, age, energy intake (MJ/day), smoking habits (smoker/nonsmoker), and use of potentially microbiome-altering medication (yes/no; metformin, psycholeptics, psychoanaleptics, proton pump inhibitors, and constipation medication)

^dAdjusted for sex, age, energy intake (MJ/day), smoking habits (smoker/nonsmoker), use of potentially microbiome-altering medication (yes/no), and BMI (kg/m²; with predictors other than BMI and WC)

for 76.9% of the constrained variance, and approximately 1% of the total variance (Online Resource 4). In qualitative interpretation, we identified a division along the second axis (RDA2) where male sex, energy intake, and smoking pointed in the opposite direction of the CRC lifestyle index and medication use. In PCoA, we observed no statistically significant differences in ordination patterns between participants in the extreme index quintiles (Q5 vs. Q1; p 0.17; Online Resource 5).

In the differential abundance analysis, we detected statistically significant associations between the CRC lifestyle index and 103 bacterial species (Online Resource 6). Of these, the majority were negative, indicating a higher relative abundance of species with lower index points. At the family level, negative associations were observed primarily with species within Lachnospiraceae, as well as with species from, for example, Coriobacteriaceae, Megasphaeraceae, and Erysipelotrichaceae. At the genus level, the negatively associated species represented, for example, *Dorea* (Lachnospiraceae), *Mediterraneibacter* (Lachnospiraceae), *Colinsella* (Coriobacteriaceae), and *Megasphaera* (Megasphaeraceae). We observed positive associations between the index and genera *Bifidobacterium* (Bifidobacteriaceae; three species) and *Haemophilus* (Pasteurellaceae; three species),

as well as with several species with incomplete taxonomy assignments across hierarchical levels.

In the cluster analysis, the bacterial species significantly associated with the CRC lifestyle index were divided into 13 clusters. Of these, five were positively and five negatively associated with the index ($p < 0.05$). Three clusters had a nonsignificant association, including a cluster formed of the three *Bifidobacterium* species ($p > 0.05$; Fig. 2, Online Resource 6). The positively associated clusters were highly diverse in terms of the included species, many of which were incompletely annotated. The negatively associated clusters were also diverse, although species in the family Lachnospiraceae were particularly prevalent. For example, cluster 3 included 18 species, of which majority were from Lachnospiraceae, such as *Mediterraneibacter* (three species), *Dorea* (two species), and *Lachnoclostridium* (one species).

Discussion

This is the first study to explore associations between higher- and lower-risk lifestyles for CRC and the gut microbiome diversity and composition. CRC-related lifestyles were modelled using a CRC lifestyle index that demonstrated good applicability in Finnish adults. The index was associated

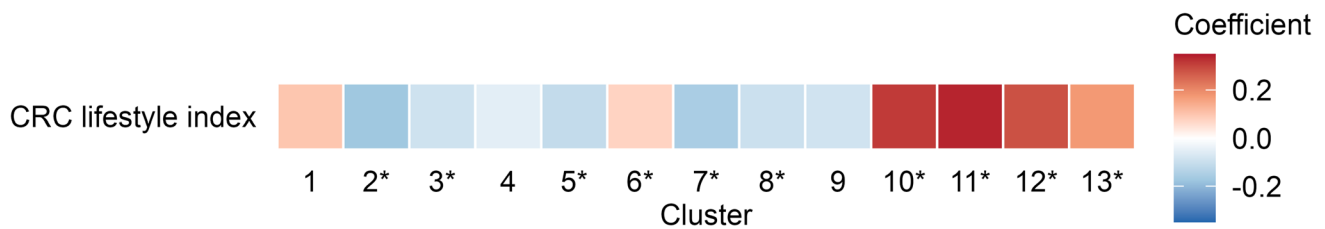


Fig. 2 Associations between the CRC lifestyle index and 13 clusters composed of the species significantly associated with the index (β coefficient). An asterisk indicates a statistically significant association between the index and a cluster ($p < 0.05$). The clusters (C) included, for example, C1: 3 species from *Bifidobacterium*; C2: 11 species from, e.g., *Selenomonas* and *Megasphaera*; C3: 18 species from, e.g., *Dorea*, *Mediterraneibacter*, *Eubacterium*, and *Lachnoclostridium*; C4: 6 species from, e.g., *Megasphaera* and *Peptostrep-*

tococcus; C5: 6 species from, e.g., *Mediterraneibacter*; C6: 8 species from, e.g., *Eubacterium*; C7: 6 species from, e.g., *Collinsella*; C8: 9 species from, e.g., *Porphyromonas*; C9: 10 species from, e.g., *Hungatella*, *Actinomyces*, *Lachnoclostridium*, and *Eubacterium*; C10: 3 species from *Haemophilus*; C11: 6 species, e.g., *Phili1 sp002069725*; C12: 5 species from *Cryptobacteroides*; C13: 12 species, e.g., *CAG-533 sp000434495*. See Online Resource 6 for further details on the clusters and species they comprise

with compositional differences in the gut microbiome between participants. Higher-risk lifestyles (lower index points) were associated with lower individual microbial diversity, higher relative abundance of bacterial species particularly in the family Lachnospiraceae, and lower relative abundance of species in the genus *Bifidobacterium*.

The CRC lifestyle index was adapted from the standardized 2018 WCRF/AICR Score [19] including nine major risk factors for CRC based on the 2018 WCRF/AICR Third Expert Report [20]. Previous prospective studies in the UK ($n = 94,000$; 863 CRC cases) [59], Spain ($n = 7,216$; 96 CRC cases) [60], and the US ($n = 45,442$ men; 1,151 CRC cases) [61] have reported a 21–48% lower CRC risk in participants with the highest versus lowest points in the standardized score or a similar index. Thus, our result of a 31% lower risk in Q5 versus Q1 fell within the range of these results, likely due to shared components across indices (BMI, red and processed meat consumption, alcohol intake). In general, the findings in prior studies corroborate our result, demonstrating that the CRC lifestyle index was an appropriate method to separate participants with higher- and lower-risk lifestyles for CRC in Finnish adults.

Despite growing interest in the gut microbiome's role in CRC pathogenesis, research on the associations between CRC-related lifestyles and microbiome diversity is limited, and evidence on individual CRC risk factors in relation to alpha or beta diversity remains inconclusive. In a systematic review comparing participants with and without obesity, results on alpha diversity were varying with most studies reporting either nonsignificant association or lower diversity in participants with obesity ($n = 20$ –1,674) [62]. BMI was, however, more consistently associated with beta diversity. Evidence on height is limited, although a positive association was observed with alpha diversity in an observational study of 3,409 US individuals participating in a commercial lifestyle intervention program [63]. Regarding physical activity, studies in a recent systematic review ($n = 49$ –868)

reported mainly nonsignificant differences in alpha diversity between highly active and inactive participants [64]. The results on beta diversity were inconsistent. In terms of dietary factors, observational studies ($n = 222$ –1,135) have mostly reported nonsignificant associations between consumption of whole grains, dairy, red meat, processed meat, or alcohol and alpha diversity, whereas some significant associations have been observed with beta diversity [65–69].

We observed statistically significant associations between several index components and alpha and beta diversity. Although our results partly aligned with the previous findings, direct comparison is limited by differences in methods and study settings; for example, diverging dietary assessment methods and sample sizes may explain discrepancies. Furthermore, most of the above-mentioned studies relied on 16S ribosomal RNA (rRNA) instead of metagenomic sequencing [62–69]. Overall, the evidence of associations between individual CRC risk factors or whole lifestyles and microbial diversity is limited and inconsistent, highlighting the need for further research.

In the differential abundance analysis, we observed a higher relative abundance of species predominantly from the family Lachnospiraceae among those with lower index points (higher-risk lifestyles). Within Lachnospiraceae, these associations were apparent in genera such as *Dorea* (two species), *Lachnoclostridium* (two species) and *Mediterraneibacter* (four species), which have all been associated with CRC in previous studies [69–71]. Of these, the two *Dorea*, one *Lachnoclostridium*, and three *Mediterraneibacter* species were also assigned to the same cluster (cluster 3), indicating similar abundance patterns and co-occurrence among them. Additionally, *Dorea* has previously been linked to obesity [62], which suggests that BMI (and WC) contributed to its negative association with the CRC index. At the species level, we identified associations with, for example, *Dorea longicatena* and *Mediterraneibacter torques*, which have been found enriched in CRC patients compared

with healthy controls in previous research [71, 72]. Within Lachnospiraceae, we also observed a negative association between the index and *Hungatella hathewayi*. *Hungatella hathewayi* has been linked with the key enzyme in the CRC-related pathway that produces trimethylamine (TMA) from choline, which is abundant in red meat and dairy [73].

Besides Lachnospiraceae, several species within families such as Erysipelotrichaceae, Coriobacteriaceae, and Megaspheeraceae, were negatively associated with the index. In a cross-sectional study of 102 Chinese adults (46 CRC patients and 56 healthy controls), both Erysipelotrichaceae and Coriobacteriaceae were enriched in CRC patients compared with healthy controls [74]. Within Coriobacteriaceae, particularly genus *Collinsella* has been linked to CRC [70, 74]. We identified four *Collinsella* species (*C. intestinalis*, *C. stercoris*, *C. ihuae*, and *C. phocaeensis*) that were negatively associated with the index. These four, along with two other Coriobacteriaceae species (*Enorma sp000333815* and *Limicola sp002160065*), were also clustered together (cluster 7). Furthermore, Coriobacteriaceae, Erysipelotrichaceae, and the genus *Megasphaera* within Megaspheeraceae have previously been linked to obesity [62, 74].

We observed positive associations between the CRC lifestyle index and three *Bifidobacterium* species (*B. longum*, *B. saguini*, and *B. breve*). In previous studies, a reduced abundance of *Bifidobacterium* has been associated with CRC particularly in the early disease stages [72, 74]. *Bifidobacterium* has also been negatively associated with obesity [75]. In contrast, high intake of dietary fiber appears to increase the abundance of *Bifidobacterium*, which is likely driven by *Bifidobacterium* being one of the genera that ferments fiber into short-chain fatty acids, SCFAs [76]. SCFAs, in turn, have several properties related, for example, to the regulation of inflammation and epigenetic mechanisms, which may benefit the prevention of colorectal carcinogenesis [3].

Overall, our results suggest that several bacterial genera and species that have been previously linked to CRC are also associated with high-risk lifestyles for CRC. Nevertheless, we did not observe significant associations with the species most consistently linked to CRC, such as *Fusobacterium nucleatum* and enterotoxigenic *Bacteroides fragilis* [4, 5]. This may be due to variation in microbial composition over the progression of carcinogenesis. For example, in a study of 616 Japanese adults, *F. nucleatum* was linked particularly with advanced CRC stages [72]. In a study of 500 German adults, *F. nucleatum* was associated with CRC but neither with pre-cancerous lesions in the gut nor with diet or other lifestyle factors [77]. As our study sample included apparently healthy adults with no CRC diagnosis at baseline, and only one incident case over the first 8 years of follow-up, our results provide information on microbial traits that potentially characterize pre-cancerous stages of CRC development in high-risk individuals. While our results support

previous research linking several microbial taxa to CRC, further research is needed to better understand the shifts in the gut microbiome throughout CRC progression, from pre-cancerous to more advanced stages. Furthermore, the functional characterization of the species associated with higher-risk lifestyles for CRC remains an important direction for future research.

Our results should be interpreted considering the study limitations. First, lifestyle data were self-reported, introducing potential bias, especially in dietary intake due to misreporting. This typically attenuates associations between dietary intake and health outcomes. In addition, different dietary assessment methods were used in testing the index (FFQ) and analyzing associations with the microbiome (48-h recall), which may have caused some discrepancy in the index performance. However, because we focused on commonly consumed foods in Finland, the recall data likely reasonably reflected participants' habitual dietary intake [23, 26, 78]. Diet was also only one of the index components. Another consideration related to the index is that, although including whole grains over dietary fiber in the index was well-justified for our study population, fiber intake may be a more relevant factor in other populations where whole grain intake is lower. Regarding the study population, in the pooled cohorts, most participants were men, which may limit the generalizability of those findings regarding the index performance overall and with respect to the microbiome analyses. However, we observed no between-cohort heterogeneity, and the results in the individual cohorts were parallel between sexes. Furthermore, in FINRISK/FINDIET 2002 used in the microbiome analyses, the study population was well balanced between men (46% of the study population) and women.

In the microbiome analyses, we included individuals who donated a fecal sample and participated in the FINDIET sub-study. These participants were likely more health-conscious compared to non-participants. Moreover, prior findings from FINRISK 2002 suggested that the sample donors were on average older, more likely women, less likely smokers, and had healthier diets than non-donors [79]. Thus, our study population may not fully represent the general Finnish adult population. However, this bias is likely reduced by the reasonably high participation rate in FINRISK/FINDIET 2002 (64%). Regarding the adjustment for constipation medication and proton-pump inhibitors, we could only account for prescription-based use, for which non-prescription use may have been overlooked. The limitations of shallow sequencing, compared with deep sequencing, in capturing genetic features and low-abundance taxa should also be acknowledged. However, we anticipate that these limitations are unlikely to substantially affect the alpha and beta diversity analyses and findings, or the differential abundance of species-level taxa presented here [42]. Overall, our results

should be primarily interpreted as exploratory and descriptive rather than predictive. For example, although statistically significant, the index and its components explained only very small proportions of the between-individual variation in microbial composition. Such modest effect sizes are typical in microbiome research and reflect the many environmental and host factors that shape microbial communities, but their predictive or clinical relevance is likely limited. Finally, the cross-sectional study design limits our ability to draw conclusions about causality.

This study also has several strengths. We tested the CRC lifestyle index in a large, pooled data of five Finnish cohorts with long median follow-up and extensive information on participants' lifestyle and health, including a validated FFQ. Information on CRC cases was obtained from the Finnish Cancer Registry, which has maintained comprehensive nationwide records of all cancer diagnoses in Finland since 1953 [40]. In the microbiome analyses, we used a large population-based study sample with comprehensive lifestyle and health data, which allowed adjustment for key confounders, including several types of medication. Furthermore, we excluded participants with a recent purchase of systemic antibacterial medication. Finally, the microbiome data were based on whole-metagenome sequencing, which provides more accurate taxonomic identification and classification than 16S rRNA gene amplicon-based sequencing [42].

To conclude, our results suggested that the diversity and composition of the gut microbiome differ between participants with higher- and lower-risk lifestyles for CRC. Higher-risk lifestyles were associated with a higher relative abundance of several microbial species previously linked to CRC. Enrichment in these species may reflect microbial alterations contributing to the initiation or progression of CRC carcinogenesis. Overall, our findings add to the current evidence of the interplay between lifestyle, the gut microbiome, and CRC risk, providing insight into microbial traits that may characterize high-risk individuals.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10552-026-02144-1>.

Acknowledgments The authors thank the participants and staff of all cohorts in this study.

Author contributions Anne-Maria Pajari is the leader of the Leg4Life project. Satu Männistö is the leader of the work package related to human nutrition and health within the Leg4Life project. Rilla Tammi, Mirikka Maukonen, Niina E Kaartinen, Kari Koponen, Teemu Niiranen, Aki S. Havulinna, and Satu Männistö contributed to study conception and design. Demetrius Albanes, Johan G. Eriksson, Pekka Jousilahti, Seppo Koskinen, Aki S. Havulinna, Guillaume Méric, Rob Knight, Veikko Salomaa, and Satu Männistö provided essential materials. Rilla Tammi performed the statistical analyses under the guidance of Mirikka Maukonen and Kari Koponen. All authors contributed to the interpretation of results. Rilla Tammi wrote the first draft of the manuscript and had the primary responsibility for the final content. All

authors commented on previous versions of the manuscript and read and approved the final manuscript.

Funding Open Access funding provided by Finnish Institute for Health and Welfare. This research was supported by the Juho Vainio Foundation (RT; Grant Number 202400258) and Yrjö Jahnsson Foundation (RT; Grant Number 20237712). In addition, this research was part of the Leg4Life Project (Legumes for Sustainable Food System and Healthy Life) funded by the Strategic Research Council at the Research Council of Finland (Grant Numbers 327698, 327699, 352481, and 352483). The authors declare that the funders had no role in the design, analysis, or writing of this article.

Data availability The datasets analyzed during the current study will be made available upon request pending application and approval through the Findata Permit Procedure at <https://www.findata.fi/en/> or through a written application to the Finnish Institute for Health and Welfare Biobank (<https://thl.fi/en/web/thl-biobank>).

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

Ethical approval This study was conducted following the guidelines laid down in the Declaration of Helsinki. Each cohort followed the ethical standards in effect at the time of data collection. ATBC was approved by the Institutional Review Groups of both the US National Cancer Institute and THL. The other cohorts were approved by the Ethics Committee of the Hospital District of Helsinki and Uusimaa.

Informed consent Informed consent was obtained from all individual participants included in this study.

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/>.

References

1. Morgan E, Arnold M, Gini A, Lorenzoni V, Cabasag CJ, Laversanne M et al (2023) Global burden of colorectal cancer in 2020 and 2040: incidence and mortality estimates from GLOBOCAN. *Gut* 72:338–344. <https://doi.org/10.1136/gutjnl-2022-327736>
2. World Cancer Research Fund/American Institute for Cancer Research (nd) Continuous Update Project Expert Report 2018. Diet, nutrition, physical activity and colorectal cancer. World Cancer Research Fund/American Institute for Cancer Research. <https://dietandcancerreport.org>. Accessed 15 Sep 2024
3. Malcomson FC (2018) Mechanisms underlying the effects of nutrition, adiposity and physical activity on colorectal cancer risk. *Nutr Bull* 43:400–415. <https://doi.org/10.1111/mbu.12359>
4. Alhinai EA, Walton GE, Commane DM (2019) The role of the gut microbiota in colorectal cancer causation. *IJMS* 20:5295. <https://doi.org/10.3390/ijms20215295>


5. Song M, Chan AT, Sun J (2020) Influence of the gut microbiome, diet, and environment on risk of colorectal cancer. *Gastroenterology* 158:322–340. <https://doi.org/10.1053/j.gastro.2019.06.048>
6. Roos E, Seppä K, Pietiläinen O, Ryyänen H, Heikkinen S, Eriksson JG et al (2022) Pairwise association of key lifestyle factors and risk of colorectal cancer: a prospective pooled multicohort study. *Cancer Rep*. <https://doi.org/10.1002/cnr2.1612>
7. Kim J, Nath K, Schmidlin K, Schaufelberger H, Quattropani C, Vannini S et al (2023) Hierarchical contribution of individual lifestyle factors and their interactions on adenomatous and serrated polyp risk. *J Gastroenterol* 58:856–867. <https://doi.org/10.1007/s00535-023-02004-8>
8. Aleksandrova K, Pischon T, Jenab M, Bueno-de-Mesquita HB, Fedirko V, Norat T et al (2014) Combined impact of healthy lifestyle factors on colorectal cancer: a large European cohort study. *BMC Med* 12:168. <https://doi.org/10.1186/s12916-014-0168-4>
9. Wei EK, Colditz GA, Giovannucci EL, Fuchs CS, Rosner BA (2009) Cumulative risk of colon cancer up to age 70 years by risk factor status using data from the Nurses' Health Study. *Am J Epidemiol* 170:863–872. <https://doi.org/10.1093/aje/kwp210>
10. Falony G, Joossens M, Vieira-Silva S, Wang J, Darzi Y, Faust K et al (2016) Population-level analysis of gut microbiome variation. *Science* 352:560–564. <https://doi.org/10.1126/science.aad3503>
11. Rothschild D, Weissbrod O, Barkan E, Kurilshikov A, Korem T, Zeevi D et al (2018) Environment dominates over host genetics in shaping human gut microbiota. *Nature* 555:210–215. <https://doi.org/10.1038/nature25973>
12. Vangay P, Johnson AJ, Ward TL, Al-Ghalith GA, Shields-Cutler RR, Hillmann BM et al (2018) US immigration westernizes the human gut microbiome. *Cell* 175:962–972. <https://doi.org/10.1016/j.cell.2018.10.029>
13. Borodulin K, Tolonen H, Jousilahti P, Jula A, Juolevi A, Koskinen S et al (2018) Cohort profile: the National FINRISK Study. *Int J Epidemiol* 47:696–696i. <https://doi.org/10.1093/ije/dyx239>
14. Männistö S, Ovaskainen M, Valsta LM (2003) The national FINDIET 2002 study, B3/2003. National Public Health Institute, Helsinki
15. The ATBC Cancer Prevention Study Group (1994) The alpha-tocopherol, beta-carotene lung cancer prevention study: design, methods, participant characteristics, and compliance. *Ann Epidemiol* 4:1–10
16. Heistaro S (2008) Methodology report: Health 2000 Survey. National Public Health Institute, Helsinki
17. Eriksson M, Rääkkönen K, Eriksson JG (2014) Early life stress and later health outcomes—findings from the Helsinki Birth Cohort Study. *Am J Hum Biol* 26:111–116. <https://doi.org/10.1002/ajhb.22502>
18. Kontinen H, Silventoinen K, Sarlio-Lähteenkorva S, Männistö S, Haukkala A (2010) Emotional eating and physical activity self-efficacy as pathways in the association between depressive symptoms and adiposity indicators. *Am J Clin Nutr* 92:1031–1039. <https://doi.org/10.3945/ajcn.2010.29732>
19. Shams-White MM, Brockton NT, Mitrou P, Romaguera D, Brown S, Bender A et al (2019) Operationalizing the 2018 World Cancer Research Fund/American Institute for Cancer Research (WCRF/AICR) cancer prevention recommendations: a standardized scoring system. *Nutrients* 11:1572. <https://doi.org/10.3390/nu11071572>
20. World Cancer Research Fund/American Institute for Cancer Research (nd) Diet, nutrition, physical activity and cancer: a global perspective. Continuous Update Project Expert Report 2018. World Cancer Research Fund/American Institute for Cancer Research. <https://dietandcancerreport.org>
21. Aune D, Chan DSM, Lau R, Vieira R, Greenwood DC, Kampman E et al (2011) Dietary fibre, whole grains, and risk of colorectal cancer: systematic review and dose–response meta-analysis of prospective studies. *BMJ* 343:d6617. <https://doi.org/10.1136/bmj.d6617>
22. Reynolds A, Mann J, Cummings J, Winter N, Mete E, Te Morenga L (2019) Carbohydrate quality and human health: a series of systematic reviews and meta-analyses. *Lancet* 393:434–445. [https://doi.org/10.1016/S0140-6736\(18\)31809-9](https://doi.org/10.1016/S0140-6736(18)31809-9)
23. Valsta L, Kaartinen N, Tapanainen H, Männistö S, Sääksjärvi K (2018) Nutrition in Finland—the National FINDIET 2017 Survey: Report no 12/2018. Finnish Institute for Health and Welfare, Helsinki
24. Slavin J (2003) Why whole grains are protective: biological mechanisms. *Proc Nutr Soc* 62:129–134. <https://doi.org/10.1079/PNS2002221>
25. Fardet A (2010) New hypotheses for the health-protective mechanisms of whole-grain cereals: what is beyond fibre? *Nutr Res Rev* 23:65–134. <https://doi.org/10.1017/S0954422410000041>
26. Männistö S, Ovaskainen ML, Valsta LM (2003) FINRAVINTO 2002—Tutkimus—the National FINDIET 2002 Study. National Public Health Institute, Helsinki
27. World Health Organization (2000) Obesity: preventing and managing the global epidemic: Report of a WHO consultation. World Health Organization, Geneva
28. World Health Organization (2008) Waist circumference and waist–hip ratio: Report of a WHO expert consultation. World Health Organization, Geneva
29. Tammi R, Männistö S, Reinivuo H, Tapanainen H, Rautanen J, Kaartinen NE (2022) The relation of whole grain surrogate estimates and food definition to total whole grain intake in the Finnish adult population. *Eur J Nutr* 62:1821–1831. <https://doi.org/10.1007/s00394-023-03119-7>
30. Blomhoff R, Andersen R, Arnesen EK, Christensen JJ, Eneroth H, Erkkola M et al (2023) Nordic Nutrition recommendations 2023. Nordic Council of Ministers, Copenhagen
31. Ovaskainen ML, Paturi M, Reinivuo H, Hannila ML, Sinkko H, Lehtisalo J et al (2008) Accuracy in the estimation of food servings against the portions in food photographs. *Eur J Clin Nutr* 62:674–681. <https://doi.org/10.1038/sj.ejcn.1602758>
32. Pietinen P, Malila N, Virtanen M, Hartman TJ, Tangrea JA, Albanes D et al (1999) Diet and risk of colorectal cancer in a cohort of Finnish men. *Cancer Causes Control* 10:387–396. <https://doi.org/10.1023/a:1008962219408>
33. Männistö S, Virtanen M, Mikkonen T, Pietinen P (1996) Reproducibility and validity of a food frequency questionnaire in a case–control study on breast cancer. *J Clin Epidemiol* 49:401–409. [https://doi.org/10.1016/0895-4356\(95\)00551-x](https://doi.org/10.1016/0895-4356(95)00551-x)
34. Paalanen L, Männistö S, Virtanen MJ, Knekt P, Räsänen L, Mantonen J et al (2006) Validity of a food frequency questionnaire varied by age and body mass index. *J Clin Epidemiol* 59:994–1001. <https://doi.org/10.1016/j.jclinepi.2006.01.002>
35. Kaartinen NE, Tapanainen H, Valsta LM, Similä ME, Reinivuo H, Korhonen T et al (2012) Relative validity of a FFQ in measuring carbohydrate fractions, dietary glycaemic index and load: exploring the effects of subject characteristics. *Br J Nutr* 107:1367–1375. <https://doi.org/10.1017/S0007114511004296>
36. Reinivuo H, Hirvonen T, Ovaskainen ML, Korhonen T, Valsta LM (2010) Dietary survey methodology of FINDIET 2007 with a risk assessment perspective. *Public Health Nutr* 13:915–919. <https://doi.org/10.1017/S1368980010001096>
37. Borodulin K, Laatikainen T, Juolevi A, Jousilahti P (2008) Thirty-year trends of physical activity in relation to age, calendar time and birth cohort in Finnish adults. *Eur J Public Health* 18:339–344. <https://doi.org/10.1093/eurpub/ckm092>
38. Tolonen H (2016) EHES manual. Part B. Field work and procedures, 2nd ed. Finnish Institute for Health and Welfare, Helsinki
39. Maukonen M, Harald K, Kaartinen NE, Tapanainen H, Albanes D, Eriksson J et al (2023) Partial substitution of red or processed

- meat with plant-based foods and the risk of type 2 diabetes. *Sci Rep* 13:5874. <https://doi.org/10.1038/s41598-023-32859-z>
40. Leinonen MK, Miettinen J, Heikkinen S, Pitkaniemi J, Malila N (2017) Quality measures of the population-based Finnish Cancer Registry indicate sound data quality for solid malignant tumours. *Eur J Cancer* 77:31–39. <https://doi.org/10.1016/j.ejca.2017.02.017>
 41. Salosensaari A, Laitinen V, Havulinna AS, Meric G, Cheng S, Perola M et al (2021) Taxonomic signatures of cause-specific mortality risk in human gut microbiome. *Nat Commun* 12:2671. <https://doi.org/10.1038/s41467-021-22962-y>
 42. Hillmann B, Al-Ghalith GA, Shields-Cutler RR, Zhu Q, Gohl DM, Beckman KB, Knight R, Knights D (2018) Evaluating the information content of shallow shotgun metagenomics. *mSystems* 3:e00069-18. <https://doi.org/10.1128/mSystems.00069-18>
 43. Chen S, Zhou Y, Chen Y, Gu J (2018) Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34:i884–i890. <https://doi.org/10.1093/bioinformatics/bty560>
 44. Guccione C, Patel L, Tomofuji Y, McDonald D, Gonzalez A, Sepich-Poore GD et al (2025) Incomplete human reference genomes can drive false sex biases and expose patient-identifying information in metagenomic data. *Nat Commun* 16:825. <https://doi.org/10.1038/s41467-025-56077-5>
 45. McDonald D, Jiang Y, Balaban M, Cantrell K, Zhu Q, Gonzalez A et al (2024) Greengenes2 unifies microbial data in a single reference tree. *Nat Biotechnol* 42:715–718. <https://doi.org/10.1038/s41587-023-01845-1>
 46. Shannon CE (1948) A mathematical theory of communication. *Bell Syst Tech J* 27:379–423
 47. Lozupone CA, Hamady M, Kelley ST, Knight R (2007) Quantitative and qualitative β diversity measures lead to different insights into factors that structure microbial communities. *Appl Environ Microbiol* 73:1576–1585. <https://doi.org/10.1128/AEM.01996-06>
 48. Smith-Warner SA, Spiegelman D, Ritz J, Albanes D, Beeson WL, Bernstein L et al (2006) Methods for pooling results of epidemiologic studies. *Am J Epidemiol* 163:1053–1064. <https://doi.org/10.1093/aje/kwj127>
 49. Anderson MJ (2008) A new method for non-parametric multivariate analysis of variance. *Austral Ecol* 26:32–46. <https://doi.org/10.1111/j.1442-9993.2001.01070.pp.x>
 50. Legendre P, Anderson MJ (1999) Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. *Ecol Monogr* 69:1–24
 51. Lin H, Peddada SD (2024) Multigroup analysis of compositions of microbiomes with covariate adjustments and repeated measures. *Nat Methods* 21:83–91. <https://doi.org/10.1038/s41592-023-02092-7>
 52. White D, Gramacy R (2022) Mapping pruning, and graphing tree models. <https://cran.r-project.org/web/packages/maptree/maptree.pdf>
 53. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B* 57:289–300
 54. R Core Team (2020) A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. <https://www.R-project.org/>
 55. Balduzzi S, Rücker G, Schwarzer G (2019) How to perform a meta-analysis with R: a practical tutorial. *Evid Based Ment Health* 22:153–160. <https://doi.org/10.1136/ebmental-2019-300117>
 56. Borman T, Ernst F, Shetty S, Lahti L (2025) Mia: microbiome analysis. <https://microbiome.github.io/mia/>
 57. Oksanen J, Blanchet G, Friendly M, Kindt R, Legendre P, McGlinn D et al (2019) Community ecology package vegan. <https://cran.r-project.org>. <https://github.com/vegandevs/vegan>. Accessed 7 June 2025
 58. Lahti L, Shetty S et al (2017) Tools for microbiome analysis in R. <https://microbiome.github.io/tutorials/>. Accessed 8 Sep 2024
 59. Malcomson FC, Parra-Soto S, Ho FK, Lu L, Celis-Morales C, Sharp L et al (2023) Adherence to the 2018 World Cancer Research Fund (WCRF)/American Institute for Cancer Research (AICR) Cancer Prevention Recommendations and risk of 14 lifestyle-related cancers in the UK Biobank prospective cohort study. *BMC Med* 21:407. <https://doi.org/10.1186/s12916-023-03107-y>
 60. Barrubés L, Babio N, Hernández-Alonso P, Toledo E, Ramírez Sabio JB, Estruch R et al (2020) Association between the 2018 WCRF/AICR and the low-risk lifestyle scores with colorectal cancer risk in the Predimed Study. *J Clin Microbiol* 9:1215. <https://doi.org/10.3390/jcm9041215>
 61. Petimar J, Smith-Warner SA, Rosner B, Chan AT, Giovannucci EL, Tabung FK (2019) Adherence to the World Cancer Research Fund/American Institute for Cancer Research 2018 recommendations for cancer prevention and risk of colorectal cancer. *Cancer Epidemiol Biomark Prev* 28:1469–1479. <https://doi.org/10.1158/1055-9965.EPI-19-0165>
 62. Pinart M, Dötsch A, Schlicht K, Laudes M, Bouwman J, Forslund SK et al (2021) Gut microbiome composition in obese and non-obese persons: a systematic review and meta-analysis. *Nutrients* 14:12. <https://doi.org/10.3390/nu14010012>
 63. Manor O, Dai CL, Kornilov SA, Smith B, Price ND, Lovejoy JC et al (2020) Health and disease markers correlate with gut microbiome composition across thousands of people. *Nat Commun* 11:5206. <https://doi.org/10.1038/s41467-020-18871-1>
 64. Pérez-Prieto I, Plaza-Florido A, Ubago-Guisado E, Ortega FB, Altmäe S (2024) Physical activity, sedentary behavior and microbiome: a systematic review and meta-analysis. *J Sci Med Sport* 27:793–804. <https://doi.org/10.1016/j.jsams.2024.07.003>
 65. Zhang Y, Chen H, Lu M, Cai J, Lu B, Luo C et al (2022) Habitual diet pattern associations with gut microbiome diversity and composition: results from a Chinese adult cohort. *Nutrients* 14:2639. <https://doi.org/10.3390/nu14132639>
 66. Aslam H, Collier F, Davis JA, Quinn TP, O’Hely M, Pasco JA et al (2021) Gut microbiome diversity and composition are associated with habitual dairy intakes: a cross-sectional study in men. *J Nutr* 151:3400–3412. <https://doi.org/10.1093/jn/nxab252>
 67. Noh H, Jang HH, Kim G, Zouiouich S, Cho SY, Kim HJ et al (2021) Taxonomic composition and diversity of the gut microbiota in relation to habitual dietary intake in Korean adults. *Nutrients* 13:366. <https://doi.org/10.3390/nu13020366>
 68. Partula V, Mondot S, Torres MJ, Kesse-Guyot E, Deschasaux M, Assmann K et al (2019) Associations between usual diet and gut microbiota composition: results from the Milieu Intérieur Cross-Sectional Study. *Am J Clin Nutr* 109:1472–1483. <https://doi.org/10.1093/ajcn/nqz029>
 69. Zhernakova A, Kurilshikov A, Bonder MJ, Tigchelaar EF, Schirmer M, Vatanen T et al (2016) Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. *Science* 352:565–569. <https://doi.org/10.1126/science.aad3369>
 70. Kim DJ, Yang J, Seo H, Lee WH, Lee HD, Kym S et al (2020) Colorectal cancer diagnostic model utilizing metagenomic and metabolomic data of stool microbial extracellular vesicles. *Sci Rep* 10:2860. <https://doi.org/10.1038/s41598-020-59529-8>
 71. Wirbel J, Pyl PT, Kartal E, Zych K, Kashani A, Milanese A et al (2019) Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer. *Nat Med* 25:679–689. <https://doi.org/10.1038/s41591-019-0406-6>
 72. Yachida S, Mizutani S, Shiroma H, Shiba S, Nakajima T, Sakamoto T et al (2019) Metagenomic and metabolomic analyses reveal distinct stage-specific phenotypes of the gut microbiota in

- colorectal cancer. *Nat Med* 25:968–976. <https://doi.org/10.1038/s41591-019-0458-7>
73. Thomas AM, Manghi P, Asnicar F, Pasolli E, Armanini F, Zolfo M et al (2019) Metagenomic analysis of colorectal cancer datasets identifies cross-cohort microbial diagnostic signatures and a link with choline degradation. *Nat Med* 25:667–678. <https://doi.org/10.1038/s41591-019-0405-7>
74. Chen W, Liu F, Ling Z, Tong X, Xiang C (2012) Human intestinal lumen and mucosa-associated microbiota in patients with colorectal cancer. *PLoS ONE* 7:e39743. <https://doi.org/10.1371/journal.pone.0039743>
75. Ozato N, Saito S, Yamaguchi T, Katashima M, Tokuda I, Sawada K et al (2019) *Blautia* genus associated with visceral fat accumulation in adults 20–76 years of age. *NPJ Biofilms Microbiomes* 5:28. <https://doi.org/10.1038/s41522-019-0101-x>
76. So D, Whelan K, Rossi M, Morrison M, Holtmann G, Kelly JT et al (2018) Dietary fiber intervention on gut microbiota composition in healthy adults: a systematic review and meta-analysis. *Am J Clin Nutr* 107:965–983. <https://doi.org/10.1093/ajcn/nqy041>
77. Amitay EL, Werner S, Vital M, Pieper DH, Höfler D, Gierse IJ et al (2017) *Fusobacterium* and colorectal cancer: causal factor or passenger? Results from a large colorectal cancer screening study. *Carcinogenesis* 38:781–788. <https://doi.org/10.1093/carcin/bgx053>
78. Tammi R, Reinivuo H, Tapanainen H, Rautanen J, Männistö S, Kaartinen NE (2021) Whole grain intake in the Finnish adult population—results of the FINDIET 2017 Study (in Finnish): Tutkimuksesta tiiviisti 64/2021. Finnish Institute for Health and Welfare, Helsinki
79. Koponen KK, Salosensaari A, Ruuskanen MO, Havulinna AS, Männistö S, Jousilahti P et al (2021) Associations of healthy food choices with gut microbiota profiles. *Am J Clin Nutr*. <https://doi.org/10.1093/ajcn/nqab077>

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

Authors and Affiliations

Rilla Tammi^{1,2,18}  · Mirkka Maukonen^{1,3} · Niina E. Kaartinen¹ · Kari Koponen^{1,4,5} · Teemu Niiranen^{1,6,7} · Guillaume Méric^{8,9,10,11} · Demetrius Albanes¹² · Johan G. Eriksson^{3,13} · Pekka Jousilahti^{1,2} · Seppo Koskinen¹ · Anne-Maria Pajari¹⁴ · Rob Knight¹⁵ · Aki S. Havulinna^{16,17} · Veikko Salomaa^{1,7} · Satu Männistö¹

✉ Rilla Tammi
rilla.tammi@thl.fi

¹ Department of Public Health, Finnish Institute for Health and Welfare (THL), Helsinki, Finland

² Department of Public Health, Faculty of Medicine, University of Helsinki, Helsinki, Finland

³ Folkhälsan Research Center, Helsinki, Finland

⁴ Department of Bacteriology and Immunology, University of Helsinki, Helsinki, Finland

⁵ Institute of Dentistry, University of Eastern Finland, Kuopio, Finland

⁶ Department of Internal Medicine, University of Turku, Turku, Finland

⁷ Division of Medicine, Turku University Hospital, Turku, Finland

⁸ Systems Medicine & Population Health Division, Department of Life Sciences, University of Bath, Claverton Down, Bath, UK

⁹ Cambridge Baker Systems Genomics Initiative, Baker Heart and Diabetes Institute, Melbourne, Australia

¹⁰ Department of Cardiovascular Research, Translation, and Implementation, La Trobe University, Melbourne, VIC, Australia

¹¹ Department of Cardiometabolic Health, University of Melbourne, Melbourne, VIC, Australia

¹² Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health (NIH), Bethesda, MD, USA

¹³ Department of General Practice and Primary Health Care, University of Helsinki and Helsinki University Hospital, Helsinki, Finland

¹⁴ Department of Food and Nutrition, University of Helsinki, Helsinki, Finland

¹⁵ University of California San Diego, La Jolla, CA, USA

¹⁶ Institute for Molecular Medicine Finland, FIMM-HiLIFE, Helsinki, Finland

¹⁷ Department of Computing, University of Turku, Turku, Finland

¹⁸ Finnish Institute for Health and Welfare (THL), P.O. Box 30, 00271 Helsinki, Finland