

1 Long term follow-up of colorectal cancer screening
2 attendees identifies differences in
3 *Phascolarctobacterium spp.* using 16S rRNA and
4 metagenome sequencing

5 Bucher-Johannessen, C.^{1,2,3}, Birkeland, E³., Vinberg, E.², Bemanian V.⁴, Hoff, G.^{5,6}, Berstad, P.⁶,
6 Rounge, TB.^{1,2,7}

7 Affiliations:

8 1 Department of Tumour Biology, Institute of Cancer Research, Oslo University Hospital, Norway

9 2 Department of Research, Cancer Registry of Norway, Norway

10 3 Centre for Bioinformatics, Department of Informatics, University of Oslo, Norway

11 4 Department of Pathology, Akershus University Hospital, Norway

12 5 Department of Research, Telemark Hospital Skien, Norway

13 6 Section for colorectal cancer screening, Cancer Registry of Norway, Oslo University Hospital, Norway

14 7 Centre for Bioinformatics, Department of Pharmacy, University of Oslo, Norway

15

16

17 **Background:** The microbiome has been implicated in the initiation and progression of colorectal
18 cancer (CRC) in cross sectional studies. However, there is a lack of studies using prospectively
19 collected samples.

20

21 **Methods:** We analysed 144 archived faecal samples from participants in the NORwegian
22 Colorectal CAncer Prevention (NORCCAP) trial diagnosed with CRC or high-risk adenomas

23 (HRA) at screening, or who remained cancer-free during 17 years of follow-up. We performed
24 16S rRNA sequencing of all samples, and metagenome sequencing on a subset of 47 samples.
25 Differences in taxonomy and gene content between outcome groups were assessed for alpha
26 and beta diversity, and differential abundance.

27
28 **Results:** Diversity and composition analyses showed no significant differences between CRC,
29 HRA, and healthy controls. *Phascolarctobacterium succinatutens* was more abundant in CRC
30 compared to healthy controls in both the 16S and metagenome data. The abundance of
31 *Bifidobacterium* and *Lachnospiraceae spp.* were associated with time to CRC diagnosis.

32
33 **Conclusion:** Using a longitudinal study design, we identified three taxa as being potentially
34 associated with CRC. These should be the focus of further studies of microbial changes occurring
35 prior to CRC diagnosis.

36
37 Key-words: archived faecal samples, colorectal cancer screening, microbiome, 16s rRNA
38 sequencing, metagenome, long term follow-up

39 INTRODUCTION

40 CRC is the third most common cancer in men and the second in women world-wide [1, 2].
41 Symptoms are often unspecific, and many cases are detected at an advanced stage with reduced
42 prospects for curative treatment. The progression towards CRC passes through stages of
43 molecular and morphological changes from small and benign, through advanced adenoma, and
44 finally to CRC. This adenoma-carcinoma sequence is estimated to take on average between 10-
45 15 years [3]. This time window provides an opportunity to screen and potentially remove lesions
46 that have not yet developed into clinical cancer and advanced stages [3, 4]. Several randomized

47 studies have estimated that CRC screening by faecal tests reduces CRC mortality by 15-30% [5-
48 8]. However, faecal based tests are hampered by both poor sensitivity and specificity, particularly
49 for detection of CRC precursor lesions [9]. Therefore, there is a need for additional markers that
50 can be used in faecal-based screening for CRC precursor lesions.

51 Analyses of the gut microbiome composition, diversity and functional potential have demonstrated
52 that the gut microbiome of CRC patients is different from that of their healthy counterparts, making
53 it a source of potential biomarkers for CRC [10-14]. The presence of certain microbes is strongly
54 associated with CRC. The most frequently reported are *Fusobacterium nucleatum*, *Bacteroides*
55 *fragilis* and pks+ *Escherichia coli*. Proposed mechanisms for a role of the microbiome in
56 carcinogenesis include DNA damage through secretion of genotoxic compounds, induction of
57 inflammation, and activation of pro-carcinogenic signalling pathways [15, 16]. While it has been
58 shown that faecal tests in combination with microbial biomarkers are superior at separating
59 healthy controls from CRC to that of a faecal test alone [17, 18], no specific bacterial profile is
60 recognized as a biomarker for CRC. Still less is known about the role of the microbiome in the
61 early stages of carcinogenesis.

62 To identify a pre-cancerous signal in the microbiome, there is a need for studies with sample
63 collection prior to diagnosis and long-term follow-up. We performed microbiome sequencing on
64 archived stool samples collected from screening attendees from the NORCCAP trial, with 17-year
65 follow-up time after sigmoidoscopy screening. This study included both screening-detected
66 cancers and CRC precursor lesions, as well as incident post-screening cancers, and healthy
67 controls. We aimed at detecting community-wide and specific differences in the microbial profiles
68 between CRC, HRA and healthy controls.

69 METHOD

70 Study design and participants

71 Details of the NORCCAP trial have been described previously [19-21]. Briefly, NORCCAP was a
72 randomized clinical trial including 20,780 individuals offered sigmoidoscopy screening in the
73 intervention arm and was performed in 1999-2000 (age group 55-64) and in 2001 (age group 50-
74 54). The study recruited participants directly from the population registry of the Norwegian
75 counties Oslo and Telemark. All participants were examined with flexible sigmoidoscopy, while
76 10,387 participants additionally delivered stool samples for an immunochemical faecal occult blood
77 test (iFOBT – FlexSure OBT) and a fresh-frozen stool sample for biobanking. We selected a
78 subset (n=300) of participants with archived fresh-frozen faecal samples for microbiome analyses
79 (Fig. 1). Participants' full CRC history was retrieved from the Cancer Registry of Norway in 2015
80 by using personal identification numbers and included the ICD-10 coded diagnoses C18, C19,
81 and C20. Individuals with high-risk adenomas were defined as persons presenting with one or
82 more adenomas of ≥ 10 mm, with high grade dysplasia or villous components regardless of polyp
83 size, or a person with three or more adenomas regardless of size, dysplasia and villosity. The
84 control group was selected from a pool of participants with no findings (i.e., no lesions) at the
85 screening examination (including low-risk adenomas) and who remained cancer-free during
86 follow-up. Controls were selected by matching on sex, age, and examination date. This study
87 received ethical approval from the Regional Committees for Medical and Health Research Ethics
88 in South-Eastern Norway (ref: 22337).

89 DNA-extraction, library preparation and sequencing

90 Participants were asked to collect stool samples immediately after defecation at home in 20 mL
91 vials and store the samples for at most seven days in freezer (-20 °C) before screening
92 sigmoidoscopy. Samples were delivered to either of the two screening centres in Oslo or
93 Telemark at time of sigmoidoscopy screening where further storage was at -20°C . We have

94 previously demonstrated the feasibility of obtaining microbiota profiles from these archived stool
95 samples [22]. Prior to DNA extraction, samples were thawed, homogenised and mixed with
96 OMNIgene gut buffer. Extraction of DNA was carried out using the QIA Symphony automated
97 extraction system, using the QIA Symphony DSP Virus/Pathogen Midikit (Qiagen, Hilden,
98 Germany), after an off-board lysis protocol with some modifications. Each sample was lysed with
99 bead-beating: a 500 µl sample aliquot was transferred to a Lysing Matrix E tube (MP Biomedicals)
100 and mixed with 700 µl phosphate-buffered saline (PBS). The mixture was then shaken at 6.5 m/s
101 for 45 seconds. After the bead-beating, 800 µl of the sample was mixed with 1055 µl of “off-board
102 lysis buffer” (proteinase K, ATL buffer, ACL buffer and nuclease-free water) and incubated at 68°C
103 for 15 min for lysis. Nucleic acid purification was performed on the QIA Symphony extraction robot
104 using the Complex800_OBL_CR22796_ID 3489 protocol. Purified DNA was eluted in 60 µl AVE-
105 buffer (Qiagen, Hilden, Germany). DNA purity was assessed using a Nanodrop2000 (Thermo
106 Fisher Scientific, MA, USA), and the concentration was measured using a Qubit instrument
107 (Thermo Fisher Scientific, MA, USA).

108 After DNA-extraction and sample quality assessment, libraries were prepared for 16S rRNA and
109 shotgun metagenome sequencing. In total, 144 of available samples had sufficient DNA for 16S
110 rRNA sequencing. Sample amplification was carried out using 16S primers S-D-Bact-0341-b-S-
111 17 (5'CCTACGGGNGGCWGCAG'3) and S-D- Bact-0785-a-A-21
112 (5'GACTACHVGGGTATCTAATCC' 3) to amplify the V3-V4 regions [23]. Amplification was
113 performed using the Truseq (TS)-tailed 1-step amplification protocol [24] with random spacers to
114 shift the sequencing start. Paired-end 300 bp sequencing of PCR amplicons was performed on
115 the Illumina MiSeq instrument (Illumina, Inc. CA, USA) (Fig. S2A). Forty-seven of the samples had
116 sufficient DNA for additional whole genome shotgun sequencing (Fig. S2B). The metagenomes
117 provide additional taxonomical resolution and improved estimates of functional potential and were
118 used for validation of the 16S rRNA sequencing results. Samples were cleaned up and

119 concentrated using AMPure XP (Beckman Coulter, IN, USA) and normalized to a total input of 4
120 ng dsDNA. Sequencing libraries were prepared using the Riptide protocol (Twist Bioscience HQ,
121 CA, USA), and sequenced on Illumina Novaseq paired end 2 x 130 bp. The Riptide protocol
122 includes linear amplification with random primers and dideoxy nucleotide-induced self-
123 termination, thereby avoiding DNA fragmentation [25]. Sequencing was performed at FIMM
124 Technology Centre in Helsinki, Finland.

125 Bioinformatics analyses

126 Initial quality control of 16S sequencing reads included removal of short reads (<50bp) and low-
127 quality bases with average quality across four bases below 30 using Trimmomatic (v.0.35.2) [26].
128 Removal of primer sequences was done using Cutadapt (v.2020.2.0) [27] with the following
129 options: forward primer: CCTACGGGNGGCWGCAG, reverse primer:
130 GACTACHVGGGTATCTAATCC, primer error 0.1 and primer overlap 3. Fastqc and multiqc
131 analyses were performed before and after trimming to ensure high quality of data [28]. Reads
132 were imported into Qiime2 [29] and amplicon sequence variants (ASV) classification was
133 performed using the Divisive Amplicon Denoising Algorithm 2 (DADA2) [30] plugin, including
134 length trimming, merging, denoising and chimera removal. ASV classification was done using the
135 SILVA 16S rRNA database (v.132) at a 97 percent similarity threshold [31]. ASV data were filtered
136 for mitochondria and chloroplasts, and were rarefied to a depth of 9,000 reads for each sample.
137 Metagenome functional profiles were predicted from the 16S data using Phylogenetic
138 Investigation of Communities by Reconstruction of Unobserved States 2 (PICRUST2) (v.2.3.0)
139 with default settings, using rarefied count-tables as input, and mapping to MetaCyc database
140 giving pathway abundance [32].

141 Metagenome reads were trimmed using Trimmomatic (v.0.66.0) [26] with a sliding window
142 approach where reads with average quality across four bases below 30 or a read length of less
143 than 30 basepairs were discarded. Following trimming, Bowtie2 (v.2.4.2) [33] and Samtools
144 (v.1.12) [34] were used with default settings to remove reads mapping to the human genome.
145 MetaPhlan3 (v.3.0.4) was used for taxonomic classification with default parameters [35]. Percent
146 abundances generated by MetaPhlan3 were transformed into count-like tables by multiplying by
147 the number of quality-trimmed reads per sample and dividing by 100. HUMAnN3 was used to
148 profile gene families encoding microbial pathways (v.3.0.0.alpha.2), aggregating the data
149 according to MetaCyc annotations using the UniRef90 (v.201901) database [35]. Pathway
150 abundance data was corrected for sequencing depth by dividing by number of trimmed reads and
151 multiplying by 10^6 .

152 Statistical analysis

153 All statistical analyses were performed using R (v.3.5.3) and visualized using ggplot2 (v.3.3.2)[36].
154 To assess differences between CRC, HRA and the control group, statistical tests were made
155 contrasting all three groups, or by combining CRC and HRAs. Additionally, analyses were
156 performed within the CRC group, using time to diagnosis as the dependent variable. Differences
157 between the three groups were evaluated using the Chi squared test for comparisons of two
158 categorical variables, and Kruskal Wallis test (or, for two-group comparisons, Mann-Whitney U
159 test) or Spearman`s correlation for comparisons of a continuous variable with a categorical and
160 continuous variable, respectively. Statistical associations were considered significant at the
161 $p<0.05$ level.

162 Microbial diversity was measured on ASV and species level for 16S and metagenome data,
163 respectively. Alpha diversity was determined using richness, Shannon and Inverse Simpson
164 indexes. Beta diversity was calculated using Bray-Curtis dissimilarity, all as implemented in the

165 Phyloseq R package [37] (v.1.26.1). Associations between beta diversity and CRC, HRA and
166 healthy controls were evaluated using permutational analysis of variance (PERMANOVA) with
167 999 permutations after adjustment for participant sex and screening centre, as implemented in
168 the adonis function of the R package vegan (v.2.5-7).

169 Differential abundance analyses were performed independently on ASV/species, genus, phylum
170 and pathways, and were adjusted for sex and screening centre. Before differential abundance
171 analyses, we applied low abundance filtering, retaining all taxa/pathways with a read count of at
172 least 10 in at least 10% of samples. Differential abundance analyses were performed using
173 negative binomial model-based Wald test implemented in the DeSeq2 package with the type
174 (poscounts) to account for the sparsity of microbiome data, and p-values were FDR adjusted to
175 control for multiple testing (v.1.22.2) [38].

176 RESULTS

177 Study population

178
179 Stool samples from 144 NORCCAP screening participants were selected for 16S sequencing
180 based on registry follow-up data and initial screening results. Metagenome sequencing was also
181 performed on 47 of these with the highest DNA amounts. All 144 participants in this study
182 underwent sigmoidoscopy. Five cases of CRC (3.5%) were detected during screening. Based on
183 registry follow-up, 23 (16%) participants received a CRC diagnosis within 17 years after screening
184 (Fig. 2, Table 1). The median time from screening to CRC diagnosis was 7.4 years (range 0-16
185 years) and the median age at CRC diagnosis was 65.7 years (range 54-77), including both
186 screening-detected and follow-up diagnosed CRC. Other screening-detected lesions included 63
187 HRAs (44% of study participants). Fifty-three (37%) participants had no findings of adenomas or

188 CRC during sigmoidoscopy and were cancer free during follow-up; these constituted the control
 189 group. The median age for all groups at sample collection was 57 years (range 51-65). We
 190 observed a significantly different distribution of sex and screening centre between CRC, HRA and
 191 healthy controls ($p < 0.05$). In total, 87 (60%) samples were from male participants and 89 (62%)
 192 samples were from the Telemark screening centre.

193 Table 1. Characteristics of study participants and samples

| Variables | 16S (n=144) | | | Metagenome (n=47) | | |
|-----------------------------------|----------------------------|---------------------------|----------------------------|-----------------------------------|-----------------------------------|-------------------------------------|
| | Control | HRA | CRC | Control | HRA | CRC |
| Men (%) | 36 (25) | 41 (28.5) | 10 (6.9) | 15 (31.9) | 12 (25.5) | 2 (4.3) |
| Women (%) | 17 (11.8) | 22 (15.3) | 18 (12.5) | 7 (14.9) | 6 (12.7) | 5 (10.6) |
| Telemark (%) | 39 (27) | 37 (25.7) | 13 (9) | 16 (34) | 11 (23.4) | 5 (10.6) |
| Oslo (%) | 14 (9.7) | 26 (18) | 15 (10.5) | 6 (12.8) | 7 (14.9) | 2 (4.3) |
| Age at sampling, median (range) | 57 (51-64) | 57 (51-64) | 60.5 (51-65) | 57 (54-64) | 58 (53-64) | 61 (55-65) |
| Age at diagnoses, median (range) | - | - | 65.7 (54-77) | - | - | 65.8 (61.1-74.3) |
| Time to diagnosis, median (range) | - | - | 7.4 (0-16) | - | - | 4.8 (0-14) |
| Reads, median (range) | 61,184 (10,261-416,286) | 48,014 (5,163-493,315) | 48,314 (23,701-510,589) | 7,356,487 (757,832-16,480,674) | 5,146,988 (998,978-20,095,193) | 7,248,142 (1,757,061-15,370,089) |
| Excluded | 0 | 2 [‡] * | 0 | 0 | 1* | 0 |

194 [‡] one person was excluded from 16S diversity analyses due to rarefaction criterion of at least 9000 reads * one
 195 person was excluded from all analyses (in both 16S and metagenome dataset) because of an E. coli infection

196

197 Gut microbiome diversity

198 16S sequencing of 144 samples generated 11.8 million trimmed reads with a median read depth
199 per sample of 50,205 (range 5,163-510,589). We identified in total 7,228 ASVs mapped to 337
200 species, 229 genus and 18 phyla. The median number of observed ASVs was 213.5 (range 79-
201 603). Metagenomic sequencing of 47 samples resulted in 361million trimmed reads with a median
202 read depth of 6.2 million reads (0.76-20.1). In total, 561 taxa were identified, including 323
203 species, 116 genus and 8 phyla. The median number of species per sample was 73 (34-107).
204 ASV distribution for individual samples showed one sample with 83% of reads belonging to two
205 ASVs within the genus *Escherichia-Shigella*. This was confirmed in the metagenome data where
206 95% of reads belonged to the species *Escherichia coli*. As this indicated an unrelated acute
207 infection, the sample was excluded from further analyses (Fig. S1).

208
209 Rarefying 16S data to 9,000 reads resulted in exclusion of one sample with lower sequencing
210 coverage, leaving 142 samples for 16S diversity analyses. Forty-six samples were used for
211 metagenome diversity analyses. We found no significant differences in alpha (unadjusted) or beta
212 diversity of taxa or pathways between CRC, HRA and healthy controls (Fig. 3A-D, 4A-D, $p>0.05$
213 for all comparisons). This finding remained consistent when grouping CRC and HRA cases
214 together, when looking at time to diagnosis, when considering metagenome data, and when
215 adjusting for sex and screening centre.

216 Differentially abundant taxa and pathways

217 We evaluated differences in abundance of individual taxa and pathways between the outcome
218 groups using the abundance of ASV/species, genus, phylum or pathways. We further assessed
219 associations of ASVs with the time elapsed from sample collection to CRC diagnosis.

220 CRC vs. Control

221 For the 16S data, the ASV *Phascolarctobacterium uncultured bacterium* and the phylum
222 *Firmicutes* were significantly more abundant in CRC than controls (FDR $p < 0.05$, Table 2 and Fig.
223 5A). Similarly, in the metagenome data *Phascolarctobacterium succinatutens* was significantly
224 more abundant in CRC. For the metagenome data, in total 9 species were differentially abundant
225 (FDR $p < 0.05$, Table 2 and Fig. 5C). Five of these were significantly higher in CRC compared to
226 control, whereas 4 were significantly lower. The genus *Acidaminococcus* was significantly higher
227 in CRC. Four pathways were significantly lower in CRC compared to controls.

228 HRA vs. Control

229 For 16S data, the genera *Azospirillum sp. 47_25* and *Escherichia-Shigella* were lower in HRA
230 compared to controls (FDR $p < 0.05$, Table 2 and Fig. 5A). The phyla *Proteobacteria* and
231 *Firmicutes* were lower and higher in HRA compared to controls, respectively. The direction of
232 differences for these phyla was similar in the metagenome data, though not significant. Twenty
233 pathways were lower in HRA based on 16S data. Of these, three pathways were related to heme
234 biosynthesis: HEMESYN2-PWY (heme biosynthesis II (anaerobic)), PWY-5920 (superpathway of
235 heme biosynthesis from glycine), and PWY0-1415 (superpathway of heme biosynthesis from
236 uroporphyrinogen-III) (FDR $p < 0.05$, Table 2 and Fig. 5B). The direction was similar for PWY0-
237 1415 in the metagenome data. We also observed differences in REDCITCYC (TCA cycle VIII
238 (helicobacter)) and the closely related pathways PWY0-42 (methylcitrate cycle I), PWY-5747
239 (methylcitrate cycle II) and GLYOXYLATE-BYPASS (glyoxylate cycle). For metagenome data the
240 species *Clostridium saccharolyticum* was significantly higher and genus *Parasutterella* was
241 significantly lower in HRA compared to controls (FDR $p < 0.05$, Table 2 and Fig. 5C).

242 HRA&CRC vs. Control

243 For 16S, when considering HRA and CRC as one group and comparing it to controls, the phylum
244 *Firmicutes* was significantly higher in HRA/CRC (FDR $p < 0.05$, Table 2 and Fig. 5A). The same
245 non-significant trend was observed in the metagenome data. Pathways CENTFERM-PWY
246 (pyruvate fermentation to butanoate) and PWY-6590 (superpathway of *Clostridium*
247 *acetobutylicum* acidogenic fermentation) were lower in HRA/CRC (FDR $p < 0.05$, Table 2 and Fig.
248 5B). For metagenome data, the species *Clostridium saccharolyticum* was significantly more
249 abundant in the HRA/CRC group (FDR $p < 0.05$, Table 2 and Fig. 5C).

250 Time to diagnosis

251 Assessing the CRC group only, those with a longer interval between sample collection and
252 diagnosis had higher abundance of one genus, *Bifidobacterium* and one ASV within the
253 *Lachnospiraceae* family. Additionally, three ASVs within *Lachnospiraceae* were lower in those
254 with a long time to diagnosis (FDR $p < 0.05$, Fig. 6 and Table 2).

255 DISCUSSION

256 Using both 16S rRNA and metagenome sequencing data, we analysed the microbial differences
257 between CRC, HRA and healthy controls of 144 screening attendees with long-term follow-up
258 data. *Phascolarctobacterium spp.*, were more abundant in the CRC compared to controls and
259 four ASVs belonging to the *Lachnospiraceae* family, and *Bifidobacterium* were associated with
260 time to CRC diagnosis. Several heme biosynthesis pathways were less abundant in HRA. We did
261 not observe compositional differences between CRC, HRA and healthy controls, and identified
262 no correlation between richness and time to diagnosis in the CRC group.

263 We identified *Phascolarctobacterium uncultured bacterium* and *Phascolarctobacterium*
264 *succinatutens* in the 16S and metagenome data respectively, as being significantly higher in CRC
265 compared to healthy controls. These annotations likely represent the same species. Three studies
266 have reported similar findings [39-41]. Interestingly, [41] et al. found an elevation in *P.*
267 *succinatutens* in the early stages of CRC, from polypoid adenomas to stage 1 CRC. *P.*
268 *succinatutens* is broadly distributed in the GI tract and converts succinate into propionate. The
269 strain can likely not ferment any other short-chain fatty acids or carbohydrates [42]. Succinate is
270 a tricarboxylic acid (TCA) cycle intermediate and is produced both by the host and the microbiota,
271 including CRC-associated bacteria *B. fragilis* and *F. nucleatum*. Increased succinate in the colon
272 has been linked to gut inflammation and disease, while increased propionate is thought to be anti-
273 inflammatory [43, 44]. Succinate is proposed to mediate cross-talk as a signaling metabolite that
274 act as a positive regulator of intestinal gluconeogenesis [44, 45] and thermogenesis [46] We also
275 report several pathways related to the TCA-cycle to be lower in HRA compared to controls. [47]
276 et al found this pathway to be increased in cancer.

277 Three pathways related to heme biosynthesis were significantly lower in the HRA group compared
278 to controls. While heme uptake, biosynthesis and export in bacteria are not fully understood [48,
279 49] , bleeding tumours release heme into the gut lumen. This might create a niche for heme
280 scavenging bacteria which could outcompete those who rely on heme biosynthesis.

281 *Bifidobacterium* and four ASVs belonging to *Lachnospiraceae* family were associated with
282 time to diagnosis. *Bifidobacterium* is a lactic acid producing bacteria, aiding in colonocyte renewal
283 and inhibiting growth of pathogens. Two studies found *Bifidobacterium* to be lower in persons with
284 lesions compared to controls [3, 50]. This is in line with our findings that lower levels are
285 associated with a shorter time to diagnosis. We observed different members of the
286 *Lachnospiraceae* family showing diverging associations with time to diagnosis. This family
287 was found to be enriched in controls compared to patients with lesions [51]. Some members of

288 the *Lachnospiraceae* family can produce the short-chain fatty acid butyrate [52]. Butyrate aids
289 in cell renewal of colonocytes, serves as a carbon source for the TCA cycle, and has anti-
290 inflammatory and anti-tumorigenic properties [53, 54].

291 We found no difference in diversity or composition between CRC, HRA and controls. Results from
292 similar studies seem to be conflicting, both for diversity and composition analyses. [11, 55-58].
293 Smaller differences in the microbiome of adenomas and healthy controls have been observed
294 than those observed between cancers and healthy controls [3, 11]. Unlike previous studies in the
295 field, many of our samples were collected from asymptomatic subjects, years before diagnosis of
296 cancer. While our results indicate no overall difference in diversity or composition, it is possible
297 that we have been underpowered or that factors related to study design and technical challenges
298 have led us to miss any small differences in these ecological measures.

299 This study has some other noteworthy limitations. Firstly, our samples were stored for 17 years
300 and could possibly be degraded. We do know that these samples have few freeze-thaw cycles
301 [22], but they were stored without a stabilizing agent which could to some extent influence the
302 composition of faecal samples [59, 60]. Further, we lack information on important confounding
303 factors such as diet, lifestyle factors, body mass index, and antibiotics use affecting microbiome
304 composition [53, 61]. Lastly, we observed a high abundance of the phylum *Firmicutes* in our 16S
305 data, but a similar composition was not observed for the metagenome data. This is likely due to
306 the choice of primers, where for marker gene studies, certain primers favour the amplification of
307 specific taxa [62]. Still, this did likely not affect the differential abundance analyses, as the bias
308 was uniform across samples.

309 Conclusions

310 The present study is, to the best of our knowledge, the first to examine gut microbiome samples
311 collected several years prior to CRC diagnosis. We did not find any differences between diversity

312 and composition of the gut microbiome and the presence of CRC, HRA, and controls. However,
313 analyses identified several taxa and pathways that were differentially abundant. Our study found
314 that the succinate-metabolising, associated with inflammation, *Phascolarctobacterium*
315 *succinatutens* was more prevalent in individuals diagnosed with CRC than in healthy controls,
316 identified using both 16S and metagenome data. In this population-based screening setting we
317 also show that CRC-associated taxa are identifiable years prior to diagnosis of CRC.

318

319 Declarations

320 Ethics approval and consent to participate

321 This study received ethical approval from the Regional Committees for Medical and Health
322 Research Ethics in South-Eastern Norway (ref: 22337).

323 Consent for publication

324 Not applicable

325 Availability of data and materials

326 The datasets generated and/or analysed during the current study are not publicly available
327 because this is human data and individuals' privacy could be compromised but data is available
328 from the corresponding author on reasonable request.

329 Competing interests

330 The authors declare that they have no competing interests

331 Funding

332 Data analyses and writing this manuscript are a part of the PhD work of CBJ which is funded by
333 South-Eastern Norway Regional Health Authority. Lab work including DNA isolation, library
334 preparation and sequencing and also data management and project coordination is funded by the
335 Cancer Registry of Norway funds.

336 Authors' contributions

337 TBR and GH designed the study. CBJ and EB analysed the data. EV and VB performed sample
338 preparation and lab work. CBJ, EB and TBR drafted the manuscript. All authors commented and
339 approved the final manuscript.

340 Acknowledgements

341 We would like to thank Jan Inge Nordby for his contribution on sample preparation and lab work.
342 Library preparation and sequencing was performed at FIMM Technology Centre supported by
343 HiLIFE and Biocenter Finland. Especially, we would like to thank Tiina Hannunen, Harri A. Kangas
344 and Pekka J. Ellonen for the good cooperation, service and communication. We would also like
345 to acknowledge Even Sannes Riiser for his early contributions on lab coordination and
346 bioinformatics analyses. Finally, this article is a result of cooperations and scientific discussions
347 among colleagues in our research group. Therefore, we want to thank Ane Sørli Kværner,
348 Ekaterina Avershina, Paula Istvan and Maja Jacobsen.

349

350

351

352

353

354

References

- 355
356
- 357 1. Dekker E, Tanis PJ, Vleugels JLA, Kasi PM, Wallace MB: **Colorectal cancer**. *The Lancet*
358 2019, **394**(10207):1467-1480.
- 359 2. Morgan E, Arnold M, Gini A, Lorenzoni V, Cabasag CJ, Laversanne M, Vignat J, Ferlay J,
360 Murphy N, Bray F: **Global burden of colorectal cancer in 2020 and 2040: incidence**
361 **and mortality estimates from GLOBOCAN**. *Gut* 2022.
- 362 3. Feng Q, Liang S, Jia H, Stadlmayr A, Tang L, Lan Z, Zhang D, Xia H, Xu X, Jie Z *et al*:
363 **Gut microbiome development along the colorectal adenoma-carcinoma sequence**.
364 *Nat Commun* 2015, **6**:6528.
- 365 4. Brenner H, Kloor M, Pox CP: **Colorectal cancer**. *The Lancet* 2014, **383**(9927):1490-1502.
- 366 5. Chiu HM, Chen SL, Yen AM, Chiu SY, Fann JC, Lee YC, Pan SL, Wu MS, Liao CS, Chen
367 HH *et al*: **Effectiveness of fecal immunochemical testing in reducing colorectal**
368 **cancer mortality from the One Million Taiwanese Screening Program**. *Cancer* 2015,
369 **121**(18):3221-3229.
- 370 6. Giorgi Rossi P, Vicentini M, Sacchetti C, Di Felice E, Caroli S, Ferrari F, Mangone L,
371 Pezzarossi A, Roncaglia F, Campari C *et al*: **Impact of Screening Program on Incidence**
372 **of Colorectal Cancer: A Cohort Study in Italy**. *The American journal of gastroenterology*
373 2015, **110**(9):1359-1366.
- 374 7. Lindholm E, Brevinge H, Haglund E: **Survival benefit in a randomized clinical trial of**
375 **faecal occult blood screening for colorectal cancer**. *Br J Surg* 2008, **95**(8):1029-1036.
- 376 8. Shaikat A, Mongin SJ, Geisser MS, Lederle FA, Bond JH, Mandel JS, Church TR: **Long-**
377 **term mortality after screening for colorectal cancer**. *N Engl J Med* 2013,
378 **369**(12):1106-1114.
- 379 9. Ribbing Wilen H, Blom J, Hoijer J, Andersson G, Lowbeer C, Hultcrantz R: **Fecal**
380 **immunochemical test in cancer screening - colonoscopy outcome in FIT positives**
381 **and negatives**. *Scandinavian journal of gastroenterology* 2019, **54**(3):303-310.
- 382 10. Yu J, Feng Q, Wong SH, Zhang D, Liang QY, Qin Y, Tang L, Zhao H, Stenvang J, Li Y *et*
383 *al*: **Metagenomic analysis of faecal microbiome as a tool towards targeted non-**
384 **invasive biomarkers for colorectal cancer**. *Gut* 2017, **66**(1):70-78.
- 385 11. Thomas AM, Manghi P, Asnicar F, Pasolli E, Armanini F, Zolfo M, Beghini F, Manara S,
386 Karcher N, Pozzi C *et al*: **Metagenomic analysis of colorectal cancer datasets**
387 **identifies cross-cohort microbial diagnostic signatures and a link with choline**
388 **degradation**. *Nat Med* 2019, **25**(4):667-678.

- 389 12. Chen D, Jin D, Huang S, Wu J, Xu M, Liu T, Dong W, Liu X, Wang S, Zhong W *et al*:
390 **Clostridium butyricum, a butyrate-producing probiotic, inhibits intestinal tumor**
391 **development through modulating Wnt signaling and gut microbiota.** *Cancer Lett*
392 2020, **469**:456-467.
- 393 13. Rebersek M: **Gut microbiome and its role in colorectal cancer.** *BMC cancer* 2021,
394 **21**(1):1325.
- 395 14. Yang J, Li D, Yang Z, Dai W, Feng X, Liu Y, Jiang Y, Li P, Li Y, Tang B *et al*: **Establishing**
396 **high-accuracy biomarkers for colorectal cancer by comparing fecal microbiomes in**
397 **patients with healthy families.** *Gut Microbes* 2020, **11**(4):918-929.
- 398 15. Pleguezuelos-Manzano C, Puschhof J, Clevers H: **Gut Microbiota in Colorectal Cancer:**
399 **Associations, Mechanisms, and Clinical Approaches.** *Annual Review of Cancer*
400 *Biology* 2021, **6**(1).
- 401 16. Wong SH, Yu J: **Gut microbiota in colorectal cancer: mechanisms of action and**
402 **clinical applications.** *Nat Rev Gastroenterol Hepatol* 2019, **16**(11):690-704.
- 403 17. Wong SH, Kwong TNY, Chow TC, Luk AKC, Dai RZW, Nakatsu G, Lam TYT, Zhang L,
404 Wu JCY, Chan FKL *et al*: **Quantitation of faecal Fusobacterium improves faecal**
405 **immunochemical test in detecting advanced colorectal neoplasia.** *Gut* 2017,
406 **66**(8):1441-1448.
- 407 18. Yuan B, Ma B, Yu J, Meng Q, Du T, Li H, Zhu Y, Sun Z, Ma S, Song C: **Fecal Bacteria**
408 **as Non-Invasive Biomarkers for Colorectal Adenocarcinoma.** *Front Oncol* 2021,
409 **11**:664321.
- 410 19. Holme O, Loberg M, Kalager M, Bretthauer M, Hernan MA, Aas E, Eide TJ, Skovlund E,
411 Schneede J, Tveit KM *et al*: **Effect of flexible sigmoidoscopy screening on colorectal**
412 **cancer incidence and mortality: a randomized clinical trial.** *JAMA* 2014, **312**(6):606-
413 615.
- 414 20. Bretthauer M, Gondal G, Larsen K, Carlsen E, Eide TJ, Grotmol T, Skovlund E, Tveit KM,
415 Vatn MH, Hoff G: **Design, organization and management of a controlled population**
416 **screening study for detection of colorectal neoplasia: attendance rates in the**
417 **NORCCAP study (Norwegian Colorectal Cancer Prevention).** *Scandinavian journal of*
418 *gastroenterology* 2002, **37**(5):568-573.
- 419 21. Bretthauer M, Thiis-Evensen E, Huppertz-Hauss G, Gisselsson L, Grotmol T, Skovlund E,
420 Hoff GJG: **NORCCAP (Norwegian colorectal cancer prevention): a randomised trial**
421 **to assess the safety and efficacy of carbon dioxide versus air insufflation in**
422 **colonoscopy.** 2002, **50**(5):604-607.
- 423 22. Rounge TB, Meisal R, Nordby JI, Ambur OH, de Lange T, Hoff G: **Evaluating gut**
424 **microbiota profiles from archived fecal samples.** *BMC Gastroenterol* 2018, **18**(1):171.

- 425 23. Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, Glockner FO:
426 **Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-**
427 **generation sequencing-based diversity studies.** *Nucleic Acids Res* 2013, **41**(1):e1.
- 428 24. Raju SC, Lagström S, Ellonen P, de Vos WM, Eriksson JG, Weiderpass E, Rounge TB:
429 **Reproducibility and repeatability of six high-throughput 16S rDNA sequencing**
430 **protocols for microbiota profiling.** *J Microbiol Methods* 2018, **147**:76-86.
- 431 25. Siddique A, Suckow G, Bahena J, Homer N, Ordoukhanian P, Head S, Brown K: **RipTide**
432 **Ultra High-Throughput Rapid DNA Library Preparation for Next Generation**
433 **Sequencing.** *Journal of Biomolecular Techniques: JBT* 2019, **30**(Suppl):S36.
- 434 26. Bolger AM, Lohse M, Usadel B: **Trimmomatic: a flexible trimmer for Illumina sequence**
435 **data.** *Bioinformatics* 2014, **30**(15):2114-2120.
- 436 27. Martin M: **Cutadapt removes adapter sequences from high-throughput sequencing**
437 **reads.** 2011 2011, **17**(1):3 %J EMBnet.journal.
- 438 28. Ewels P, Magnusson M, Lundin S, Kaller M: **MultiQC: summarize analysis results for**
439 **multiple tools and samples in a single report.** *Bioinformatics* 2016, **32**(19):3047-3048.
- 440 29. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H,
441 Alm EJ, Arumugam M, Asnicar F *et al*: **Reproducible, interactive, scalable and**
442 **extensible microbiome data science using QIIME 2.** *Nat Biotechnol* 2019, **37**(8):852-
443 857.
- 444 30. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP: **DADA2: High-**
445 **resolution sample inference from Illumina amplicon data.** *Nat Methods* 2016,
446 **13**(7):581-583.
- 447 31. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glockner FO:
448 **The SILVA ribosomal RNA gene database project: improved data processing and**
449 **web-based tools.** *Nucleic Acids Res* 2013, **41**(Database issue):D590-596.
- 450 32. Douglas GM, Maffei VJ, Zaneveld JR, Yurgel SN, Brown JR, Taylor CM, Huttenhower C,
451 Langille MG: **PICRUSt2 for prediction of metagenome functions.** *Nat Biotechnol* 2020,
452 **38**(6):685-688.
- 453 33. Langmead B, Trapnell C, Pop M, Salzberg SL: **Ultrafast and memory-efficient**
454 **alignment of short DNA sequences to the human genome.** *Genome Biol* 2009,
455 **10**(3):R25.
- 456 34. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin
457 R, Genome Project Data Processing S: **The Sequence Alignment/Map format and**
458 **SAMtools.** *Bioinformatics* 2009, **25**(16):2078-2079.

- 459 35. Beghini F, McIver LJ, Blanco-Miguez A, Dubois L, Asnicar F, Maharjan S, Mailyan A,
460 Manghi P, Scholz M, Thomas AM *et al*: **Integrating taxonomic, functional, and strain-**
461 **level profiling of diverse microbial communities with bioBakery 3.** *Elife* 2021, **10**.
- 462 36. Wickham H: **ggplot2: elegant graphics for data analysis:** Springer; 2016.
- 463 37. McMurdie PJ, Holmes SJPo: **phyloseq: an R package for reproducible interactive**
464 **analysis and graphics of microbiome census data.** 2013, **8(4):e61217.**
- 465 38. Love MI, Huber W, Anders S: **Moderated estimation of fold change and dispersion for**
466 **RNA-seq data with DESeq2.** *Genome Biol* 2014, **15(12):550.**
- 467 39. Yang HJ, Kwon MJ, Chang Y, Song SK, Ahn KS, Kim HN, Yun Y, Kim HL, Park DI: **Fecal**
468 **Microbiota Differences According to the Risk of Advanced Colorectal Neoplasms.** *J*
469 *Clin Gastroenterol* 2019, **53(3):197-203.**
- 470 40. Zackular JP, Rogers MA, Ruffin MTt, Schloss PD: **The human gut microbiome as a**
471 **screening tool for colorectal cancer.** *Cancer Prev Res (Phila)* 2014, **7(11):1112-1121.**
- 472 41. Yachida S, Mizutani S, Shiroma H, Shiba S, Nakajima T, Sakamoto T, Watanabe H,
473 Masuda K, Nishimoto Y, Kubo M *et al*: **Metagenomic and metabolomic analyses reveal**
474 **distinct stage-specific phenotypes of the gut microbiota in colorectal cancer.** *Nat*
475 *Med* 2019, **25(6):968-976.**
- 476 42. Watanabe Y, Nagai F, Morotomi M: **Characterization of Phascolarctobacterium**
477 **succinatutens sp. nov., an asaccharolytic, succinate-utilizing bacterium isolated**
478 **from human feces.** *Appl Environ Microbiol* 2012, **78(2):511-518.**
- 479 43. Ternes D, Karta J, Tsenkova M, Wilmes P, Haan S, Letellier E: **Microbiome in Colorectal**
480 **Cancer: How to Get from Meta-omics to Mechanism?** *Trends Microbiol* 2020,
481 **28(5):401-423.**
- 482 44. Fernandez-Veledo S, Vendrell J: **Gut microbiota-derived succinate: Friend or foe in**
483 **human metabolic diseases?** *Rev Endocr Metab Disord* 2019, **20(4):439-447.**
- 484 45. De Vadder F, Kovatcheva-Datchary P, Zitoun C, Duchamp A, Backhed F, Mithieux G:
485 **Microbiota-Produced Succinate Improves Glucose Homeostasis via Intestinal**
486 **Gluconeogenesis.** *Cell Metab* 2016, **24(1):151-157.**
- 487 46. Mills EL, Pierce KA, Jedrychowski MP, Garrity R, Winther S, Vidoni S, Yoneshiro T,
488 Spinelli JB, Lu GZ, Kazak L *et al*: **Accumulation of succinate controls activation of**
489 **adipose tissue thermogenesis.** *Nature* 2018, **560(7716):102-106.**
- 490 47. Vogtman E, Hua X, Zeller G, Sunagawa S, Voigt AY, Hercog R, Goedert JJ, Shi J, Bork
491 P, Sinha R: **Colorectal Cancer and the Human Gut Microbiome: Reproducibility with**
492 **Whole-Genome Shotgun Sequencing.** *PLoS one* 2016, **11(5):e0155362.**

- 493 48. Donegan RK: **The role of host heme in bacterial infection.** *Biol Chem* 2022, **403**(11-
494 12):1017-1029.
- 495 49. Fiorito V, Chiabrando D, Petrillo S, Bertino F, Tolosano E: **The Multifaceted Role of**
496 **Heme in Cancer.** *Front Oncol* 2019, **9**:1540.
- 497 50. Rezasoltani S, Asadzadeh Aghdaei H, Dabiri H, Akhavan Sepahi A, Modarressi MH,
498 Nazemalhosseini Mojarad E: **The association between fecal microbiota and different**
499 **types of colorectal polyp as precursors of colorectal cancer.** *Microb Pathog* 2018,
500 **124**:244-249.
- 501 51. Baxter NT, Ruffin MTt, Rogers MA, Schloss PD: **Microbiota-based model improves the**
502 **sensitivity of fecal immunochemical test for detecting colonic lesions.** *Genome Med*
503 2016, **8**(1):37.
- 504 52. Meehan CJ, Beiko RG: **A phylogenomic view of ecological specialization in the**
505 **Lachnospiraceae, a family of digestive tract-associated bacteria.** *Genome Biol Evol*
506 2014, **6**(3):703-713.
- 507 53. O'Keefe SJ: **Diet, microorganisms and their metabolites, and colon cancer.** *Nat Rev*
508 *Gastroenterol Hepatol* 2016, **13**(12):691-706.
- 509 54. Pryde SE, Duncan SH, Hold GL, Stewart CS, Flint HJ: **The microbiology of butyrate**
510 **formation in the human colon.** *FEMS microbiology letters* 2002, **217**(2):133-139.
- 511 55. Mira-Pascual L, Cabrera-Rubio R, Ocon S, Costales P, Parra A, Suarez A, Moris F,
512 Rodrigo L, Mira A, Collado MC: **Microbial mucosal colonic shifts associated with the**
513 **development of colorectal cancer reveal the presence of different bacterial and**
514 **archaeal biomarkers.** *J Gastroenterol* 2015, **50**(2):167-179.
- 515 56. Ahn J, Sinha R, Pei Z, Dominianni C, Wu J, Shi J, Goedert JJ, Hayes RB, Yang L: **Human**
516 **gut microbiome and risk for colorectal cancer.** *J Natl Cancer Inst* 2013, **105**(24):1907-
517 1911.
- 518 57. Zhang H, Chang Y, Zheng Q, Zhang R, Hu C, Jia W: **Altered intestinal microbiota**
519 **associated with colorectal cancer.** *Front Med* 2019, **13**(4):461-470.
- 520 58. Wirbel J, Pyl PT, Kartal E, Zych K, Kashani A, Milanese A, Fleck JS, Voigt AY, Palleja A,
521 Ponnudurai R *et al*: **Meta-analysis of fecal metagenomes reveals global microbial**
522 **signatures that are specific for colorectal cancer.** *Nat Med* 2019, **25**(4):679-689.
- 523 59. Gorzelak MA, Gill SK, Tasnim N, Ahmadi-Vand Z, Jay M, Gibson DL: **Methods for**
524 **Improving Human Gut Microbiome Data by Reducing Variability through Sample**
525 **Processing and Storage of Stool.** *PloS one* 2015, **10**(8):e0134802.

526 60. Flores R, Shi J, Yu G, Ma B, Ravel J, Goedert JJ, Sinha R: **Collection media and delayed**
527 **freezing effects on microbial composition of human stool.** *Microbiome* 2015, **3**:33.

528 61. Vujkovic-Cvijin I, Sklar J, Jiang L, Natarajan L, Knight R, Belkaid Y: **Host variables**
529 **confound gut microbiota studies of human disease.** *Nature* 2020, **587**(7834):448-454.

530 62. Soergel DA, Dey N, Knight R, Brenner SE: **Selection of primers for optimal taxonomic**
531 **classification of environmental 16S rRNA gene sequences.** *ISME J* 2012, **6**(7):1440-
532 1444.
533

534

535

536

537

538

539

540

541

542

543

544

545

546

547

548

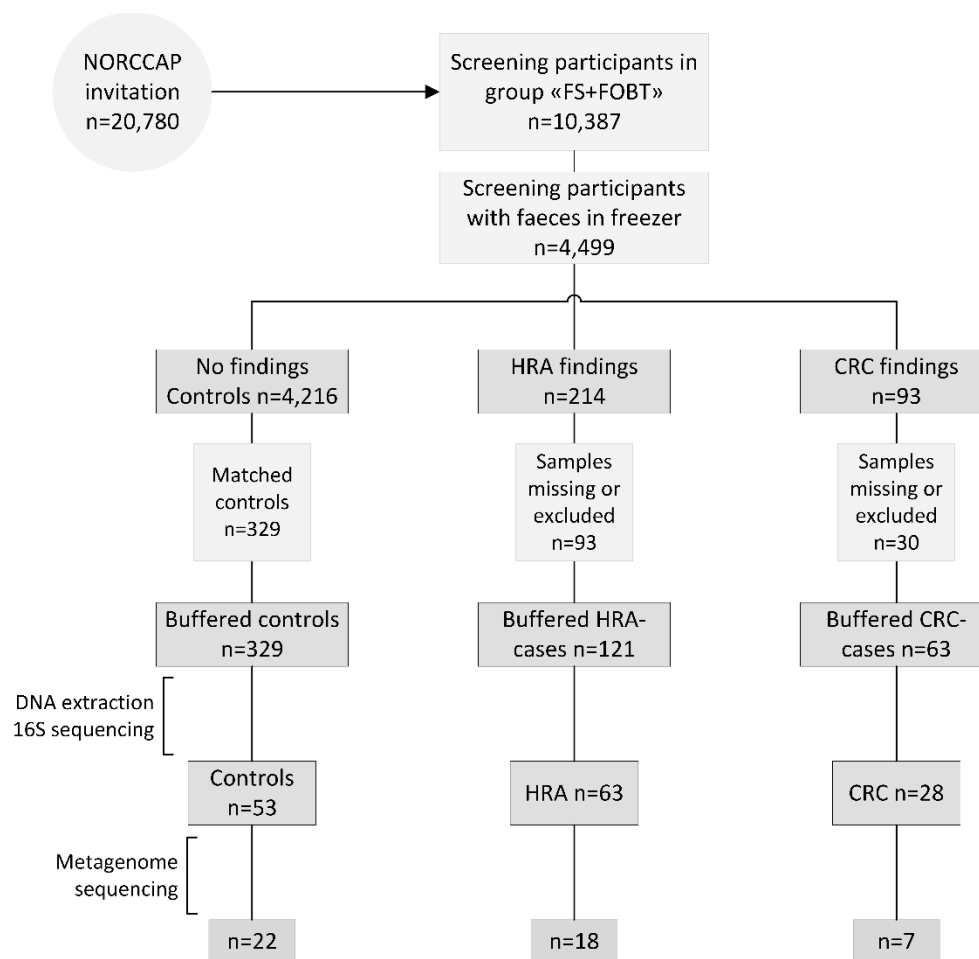
549

550

551

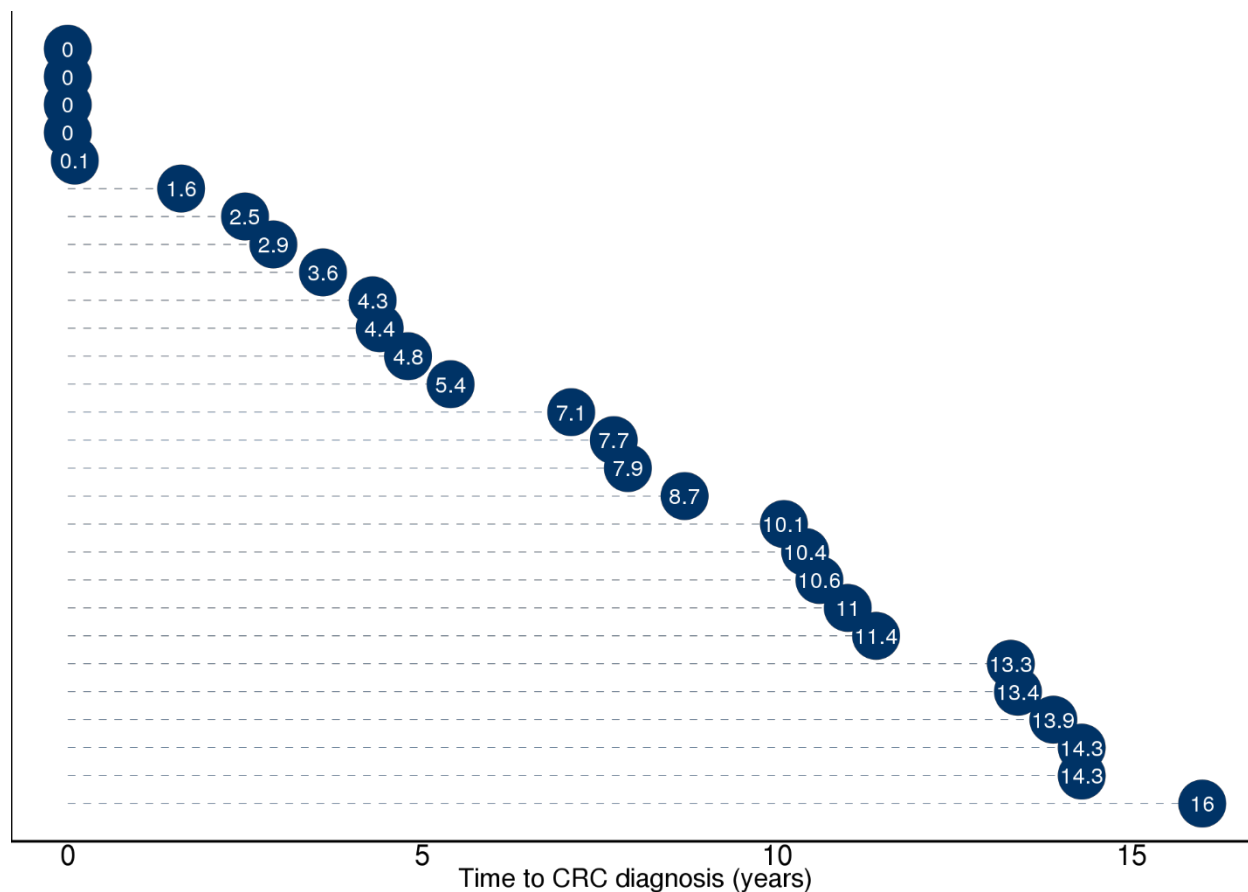
552

553



554
555 Figure 1 Recruitment flowchart. Half of the NORCCAP participants were invited to deliver a stool sample in addition
556 to participating in sigmoidoscopy screening. Half of these faecal samples were stored below -20°C. A subset of
557 samples diagnosed with CRC, HRA and matched controls were included in the study and homogenised in
558 preservation buffers. Those with sufficient DNA extracted were included in 16S rRNA (n=144) and metagenome
559 sequencing (n=47). FS fecal sigmoidoscopy, FOBT fecal occult blood test.

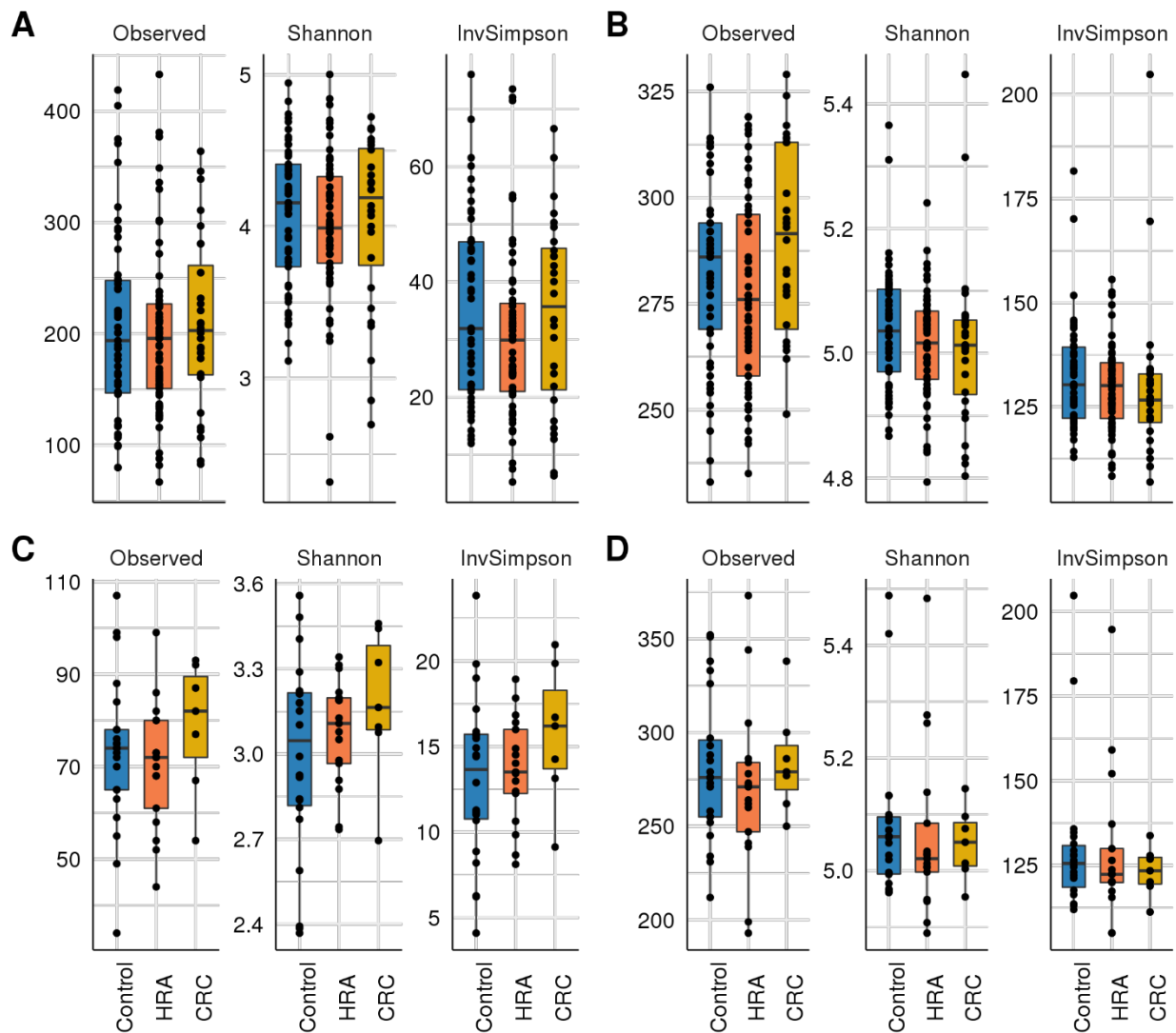
560
561



562
563 Figure 2 Overview of time from sample collection to cancer diagnosis for the 23 study participants who received a
564 CRC diagnosis during screening or follow-up. Five participants with time to diagnosis ≤ 0.1 years received the
565 diagnosis during screening.

566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586

587



588

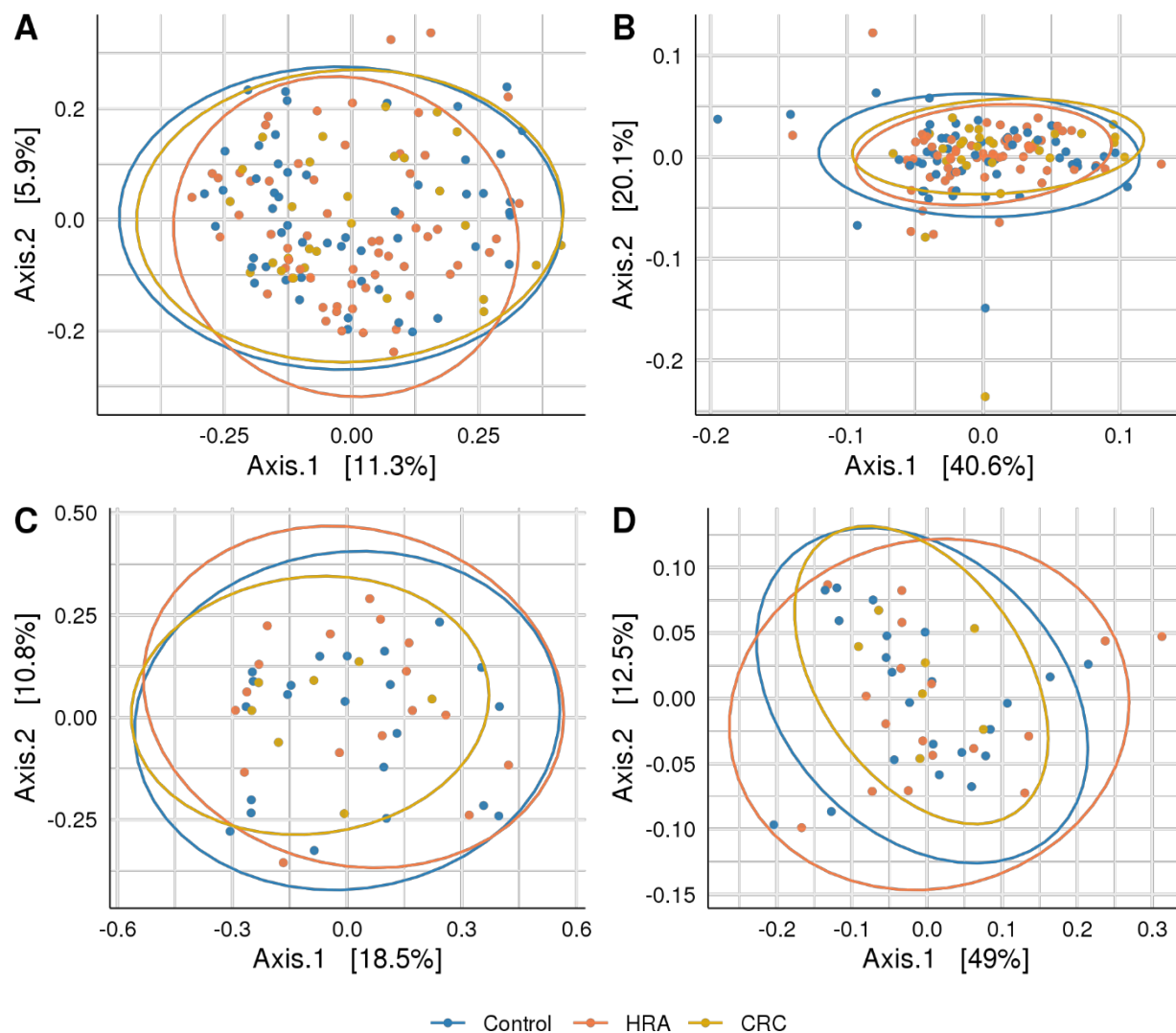
589

590

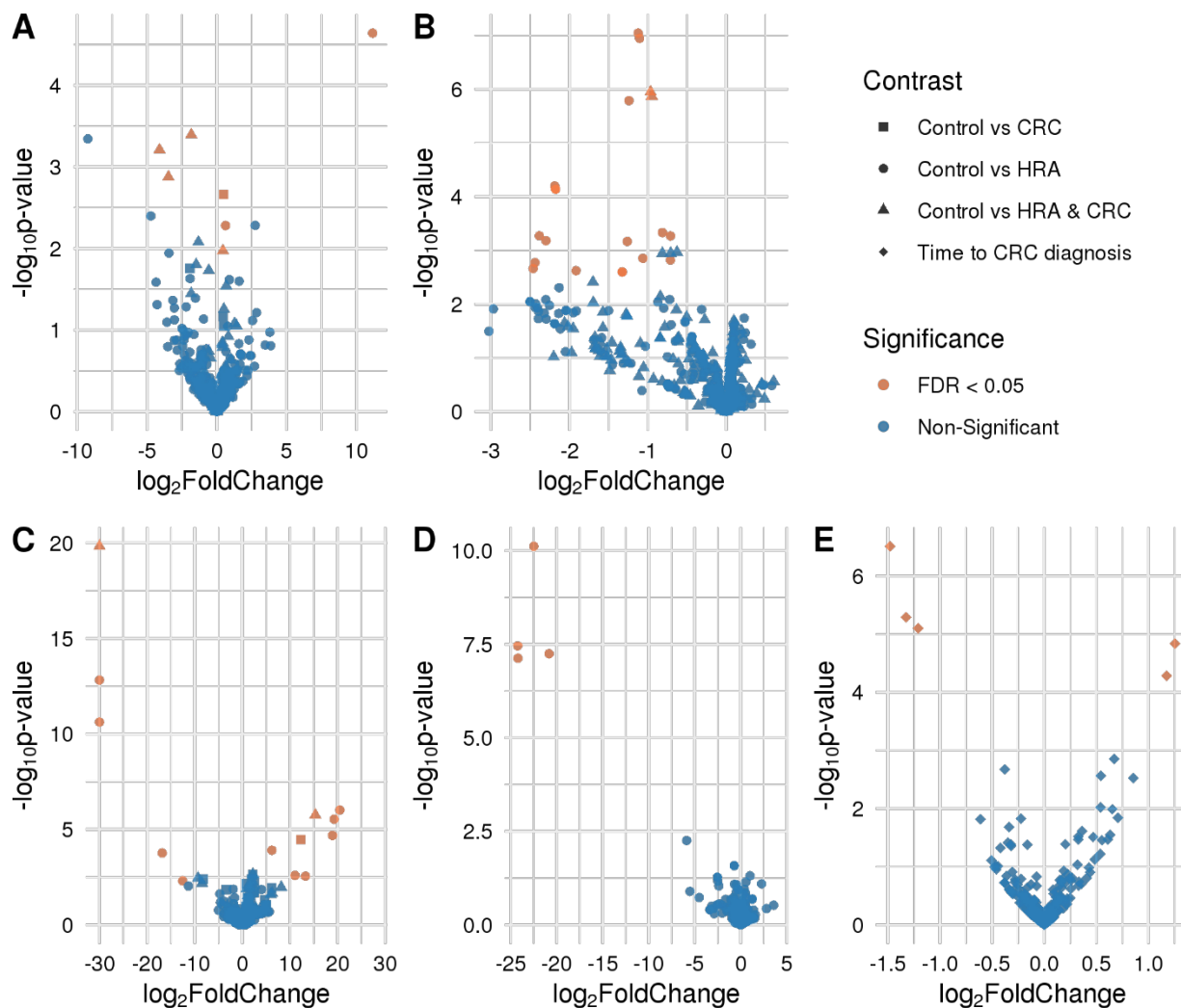
591

592

Figure 3 Alpha diversity: box plots with taxa/pathway richness (Observed), and Shannon and Inverse Simpson (InvSimpson) diversity indices in CRC, HRA and controls for A) amplicon sequence variants from 16S sequencing data B) estimated MetaCyc pathways derived from 16S data C) Species abundance based on metagenome shotgun sequencing D) Metacyc pathways based on metagenome shotgun sequencing.



593
594 Figure 4 Beta diversity: PCoA plots with Bray Curtis dissimilarity indices between CRC, HRA and controls for A)
595 amplicon sequence variants from 16S sequencing data B) MetaCyc pathways derived from 16S sequencing data C)
596 Species abundance based on metagenome shotgun sequencing D) Metacyc pathways based on metagenome shotgun
597 sequencing. Ellipses describe 95% of group variation for the principal coordinate axes.



598
 599 Figure 5 Volcano plots showing differences in abundance of taxa and pathways between groups. FDR-significant
 600 differentially abundant taxa or pathways are highlighted in red. Group comparisons are indicated by different shapes
 601 where the control group or a shorter time to diagnosis is considered the reference. Differential abundance was analysed
 602 for A) amplicon sequence variants from 16S sequencing data B) MetaCyc pathways derived from 16S sequencing data
 603 C) Species abundance based on metagenome shotgun sequencing D) Metacyc pathways based on metagenome
 604 shotgun sequencing E) amplicon sequence variants from 16S sequencing data and time to CRC diagnosis.

605
 606
 607
 608
 609
 610
 611
 612
 613
 614
 615
 616

617 Table 2 Differential abundance analyses of taxa and pathways between CRC, HRA and healthy controls. Log-2foldchange indicates the magnitude and direction of
 618 difference in abundance. Analyses were adjusted for sex and screening centre.

| Contrast | baseMean | log2FoldChange | padj | Taxa/Pathway |
|----------------------------|-----------------|-----------------------|-------------|--|
| ASV taxonomy, n=143 | | | | |
| Control vs CRC | 41,37 | 11,16 | 0,0072 | Phascolarctobacterium uncultured bacterium |
| Control vs CRC | 61410,72 | 0,63 | 0,0366 | Firmicutes |
| Control vs HRA | 33,60 | -4,10 | 0,0464 | Azospirillum sp. 47_25 |
| Control vs HRA | 303,86 | -3,47 | 0,0498 | Escherichia-Shigella |
| Control vs HRA | 662,44 | -1,83 | 0,0028 | Proteobacteria |
| Control vs HRA | 61410,72 | 0,44 | 0,0374 | Firmicutes |
| Control vs HRA/CRC | 61410,72 | 0,50 | 0,0131 | Firmicutes |
| ASV pathways, n=143 | | | | |
| Control vs HRA | 181,995 | -1,125 | 1,80E-05 | CENTFERM-PWY: pyruvate fermentation to butanoate |
| Control vs HRA | 135,426 | -1,240 | 0,0002 | FAO-PWY: fatty acid β-oxidation I |
| Control vs HRA | 48,787 | -1,325 | 0,0397 | GALACTARDEG-PWY: D-galactarate degradation I |
| Control vs HRA | 48,787 | -1,325 | 0,0397 | GLUCARGALACTSUPER-PWY: superpathway of D-glucarate and D-galactarate degradation |
| Control vs HRA | 45,812 | -2,299 | 0,0165 | GLYCOL-GLYOXDEG-PWY: superpathway of glycol metabolism and degradation |
| Control vs HRA | 60,372 | -2,383 | 0,0155 | GLYOXYLATE-BYPASS: glyoxylate cycle |
| Control vs HRA | 179,086 | -0,713 | 0,0318 | HEMESYN2-PWY: heme biosynthesis II (anaerobic) |
| Control vs HRA | 94,118 | -0,814 | 0,0155 | PWY-5177: glutaryl-CoA degradation |
| Control vs HRA | 25,662 | -2,439 | 0,0333 | PWY-5747: 2-methylcitrate cycle II |
| Control vs HRA | 58,730 | -2,176 | 0,0029 | PWY-5855: ubiquinol-7 biosynthesis (prokaryotic) |
| Control vs HRA | 58,730 | -2,176 | 0,0029 | PWY-5856: ubiquinol-9 biosynthesis (prokaryotic) |
| Control vs HRA | 58,730 | -2,176 | 0,0029 | PWY-5857: ubiquinol-10 biosynthesis (prokaryotic) |
| Control vs HRA | 29,454 | -1,915 | 0,0397 | PWY-5920: superpathway of heme biosynthesis from glycine |
| Control vs HRA | 230,973 | -1,105 | 1,80E-05 | PWY-6590: superpathway of Clostridium acetobutylicum acidogenic fermentation |
| Control vs HRA | 58,730 | -2,176 | 0,0029 | PWY-6708: ubiquinol-8 biosynthesis (prokaryotic) |
| Control vs HRA | 60,697 | -1,262 | 0,0165 | PWY0-1415: superpathway of heme biosynthesis from uroporphyrinogen-III |
| Control vs HRA | 132,568 | -1,063 | 0,0317 | PWY0-1533: methylphosphonate degradation I |
| Control vs HRA | 25,410 | -2,461 | 0,0397 | PWY0-42: 2-methylcitrate cycle I |
| Control vs HRA | 468,521 | -0,711 | 0,0155 | REDCITCYC: TCA cycle VIII (helicobacter) |

| | | | | |
|--------------------|---------|--------|--------|--|
| Control vs HRA | 56,227 | -2,187 | 0,0029 | UBISYN-PWY: superpathway of ubiquinol-8 biosynthesis (prokaryotic) |
| Control vs HRA/CRC | 181,995 | -0,965 | 0,0002 | CENTFERM-PWY: pyruvate fermentation to butanoate |
| Control vs HRA/CRC | 230,973 | -0,949 | 0,0002 | PWY-6590: superpathway of Clostridium acetobutylicum acidogenic fermentation |

Metagenome taxonomy, n=46

| | | | | |
|--------------------|----------|----------|----------|-------------------------------------|
| Control vs CRC | 3436,272 | -30 | 1,07E-09 | Bacteroides finegoldii |
| Control vs CRC | 1543,539 | 11,06101 | 3,12E-02 | Lactobacillus rogosae |
| Control vs CRC | 3173,952 | -12,5276 | 4,94E-02 | Monoglobus pectinilyticus |
| Control vs CRC | 80657,51 | 6,196732 | 2,21E-03 | Coprococcus eutactus |
| Control vs CRC | 4742,03 | -16,8273 | 2,57E-03 | Roseburia sp. CAG:303 |
| Control vs CRC | 1160,284 | 13,24881 | 3,12E-02 | Firmicutes bacterium CAG:95 |
| Control vs CRC | 3457,533 | 18,9179 | 4,59E-04 | Acidaminococcus intestini |
| Control vs CRC | 19999,34 | 20,45545 | 2,88E-05 | Phascolarctobacterium succinatutens |
| Control vs CRC | 2400,172 | -30 | 1,35E-11 | Veillonella parvula |
| Control vs CRC | 3264,922 | 19,253 | 0,0001 | Acidaminococcus |
| Control vs HRA | 511,8308 | 15,320 | 0,0002 | Clostridium saccharolyticum |
| Control vs HRA | 1584,142 | -30,000 | 7,13E-19 | Parasutterella |
| Control vs HRA/CRC | 511,8308 | 12,21184 | 0,0031 | Clostridium saccharolyticum |

Metagenome pathways, n=46

| | | | | |
|----------------|--------|---------|----------|--|
| Control vs CRC | 13,053 | -24,231 | 5,64E-06 | ENTBACSYN-PWY: enterobactin biosynthesis |
| Control vs CRC | 20,312 | -22,480 | 2,46E-08 | PWY-6285: superpathway of fatty acids biosynthesis (E. coli) |
| Control vs CRC | 5,721 | -20,822 | 6,05E-06 | PWY-6992: 1,5-anhydrofructose degradation |
| Control vs CRC | 3,859 | -24,204 | 6,05E-06 | THREOCAT-PWY: superpathway of L-threonine metabolism |

ASV taxonomy: time to CRC diagnosis, n=28

| | | | | |
|------|----------|----------|--------|-----------------|
| Time | 18,42315 | 1,251612 | 0,0008 | Bifidobacterium |
| Time | 19,47221 | -1,20951 | 0,0006 | Lachnospiraceae |
| Time | 11,94936 | -1,47904 | 0,0001 | Lachnospiraceae |
| Time | 7,238016 | -1,32495 | 0,0006 | Lachnospiraceae |
| Time | 13,97661 | 1,171725 | 0,0023 | Lachnospiraceae |