

1 **Introducing ExHiBITT – Exploring Host microbiome inTeractions in Twins-, a colon**
2 **multiomic cohort study**

3 Marina Mora-Ortiz^{1#}, Hajir Ibraheim^{2#}, Sherine Hermangild Kottoor², Ruth C. E. Bowyer¹, Sarah
4 Metrustry¹, Jeremy Sanderson³, Nicholas Powell², Tim D. Spector¹, Kerrin S. Small¹, Claire J.
5 Steves^{*1}

6 [#]Joint first authors

7 ¹ The Department of Twin Research and Genetic Epidemiology, King's College London, St
8 Thomas' Hospital, Lambeth Palace Road, London, SE1 7EH, UK;

9 ² Experimental Immunobiology, Division of Transplant Immunology and Mucosal Biology,
10 Great Maze Pond, King's College London, SE1 9RT, UK;

11 ³ Department of Gastroenterology, Guy's & St. Thomas' Hospitals NHS Foundation Trust,
12 London, United Kingdom*

13 Corresponding authors*: Claire Steves Claire.steves@kcl.ac.uk and Kerrin Small
14 Kerrin.small@kcl.ac.uk

15 **Abstract:**

16 The colon is populated by approximately 10¹² microorganisms, but the relationships between this
17 microbiome and the host health status are still not completely understood. Participants from the
18 TwinsUK cohort were recruited to study the interactions between the microbiome and host
19 adaptive immunity. In total, 205 monozygotic twins were recruited from the wider TwinsUK
20 cohort. They completed health questionnaires, and provided saliva, blood, colon biopsies from
21 three different locations, caecal fluid, and two faecal- samples.

22 Here, our objective is to present the cohort characteristics of **ExHiBITT** including i)
23 biomedical phenotypes, ii) environmental factors and ii) colonoscopic findings. A significant
24 proportion of this apparently normal cohort had colonic polyps (28%), which are of interest as
25 potential precursors of colorectal cancer, and as expected, the number of polyps found was
26 significantly correlated with BMI and age. Hitherto undiagnosed diverticulosis was also not

27 infrequently found during colonoscopy (26%) and was associated in changes in Hybrid Th1-17
28 cells in the colon. Twin proband cooccurrence rate for diverticulosis (82%), was much higher
29 than for polyps (42%). Familial factors affecting morphology or tolerance may contribute to the
30 ease of endoscopy, as both the time to reach the caecum, and pain perceived were highly
31 concordant (proband concordance: 85% and 56% respectively). We found the expected positive
32 relationship between BMI and colonoscopic anomalies such as diverticular disease and polyps in
33 the whole population, but within twin pairs this association was reversed. This suggests that
34 familial factors confound these associations. Host and microbial Next Generation Sequencing
35 and metabolomics of the samples collected are planned in this cohort.

36 **Key words:** cohort profile, twins, colon, microbiome, host genetics, polyps, diverticulosis

37 **Introduction**

38 The colon, is the last part of the digestive system where water, salt and some vitamins, such as
39 vitamin K or thiamine, are absorbed prior to defaecation. It is also a key location where
40 microbial fermentation of remaining solid waste material takes place (1-3). The large intestine is
41 populated by approximately 10^{12} microorganisms, out of the circa 10^{14} microorganisms hosted in
42 different niches of the human body including skin, genitourinary and respiratory tracts, and small
43 and large intestine (4-7). Over 700 different species live in the colon, prevalently dominated by
44 Firmicutes and Bacteroides, with a varying ratio depending on different factors including health
45 status (8-11). Interactions between the microbiota and the colon can be classified as mutualistic,
46 symbiotic or pathobiontic (12, 13) and evidence is mounting for a role in host health and disease.
47 Most human studies to date investigate the relationship between faecal samples and host
48 physiology. However, animal studies have indicated tighter relationships between colonic
49 microbiota and host physiology than with the stool, and highlighted the influence of microbiota
50 on colonic gene expression (14).

51 ExHiBITT – Exploring Host microBIome inTeration in Twins- is a sub-study within TwinsUK
52 cohort (15, 16), which will enable scientist access to a large number of OMICS' data related to

53 the colon. Twin studies may particularly useful to study deep-tissue microbiota-host interactions,
54 in part because of the strong influence of host genetics on gene expression and immune function
55 disease associations, and to a lesser extent on microbiome itself (17-20). By analysing changes in
56 monozygotic twins, with the same host genetics, effects of different microbiota can be examined
57 without the variance attributable to host genetics. Thus, twins' studies are recognised for their
58 potential to investigate different phenotypes separating genetic from environmental effects (21).
59 Monozygotic (MZ) twin pairs, where genetic variation is rare or null, provide the ideal scenario
60 to investigate the effect of environmental factors such as gut microbiota, diet, smoking status or
61 living habitat (22, 23). Analysis of samples, using high-throughput techniques, including Next
62 Generation Sequencing (NGS), metabolomics and immune profiling of peripheral blood and
63 caecum of twin pairs, is underway to investigate the host and microbiome genetics, metabolome
64 and associated modulations of the immune system.

65 The objective of the present study was firstly to describe the distribution of this newly
66 established cohort according to three different types of phenotypes: i) colonoscopy findings, ii)
67 biomedical phenotypes and iii) environmental factors, and to assess the twin concordance for
68 endoscopic variables. We then interrogated the relationships between BMI and colonoscopic
69 findings using standard and within-pair regression modelling. Secondly, although routine
70 endoscopic biopsies are considered safe, there is limited outcome data in patients that have large
71 numbers of research biopsies taken, and where available is retrospective in nature (24). In this
72 study we report on the safety and tolerability of taking more than 20 research biopsies within an
73 older adult population.

74 Analysis of the colonoscopic findings found within this non-clinical population are not trivial.
75 Polyps are tumours affecting approximately half of the western population at some point in life
76 and detected in up to a third of all colonoscopies (25). The majority of polyps are adenomatous
77 and, by definition, dysplastic with malignant potential. Adenomatous polyps increase with age,
78 occurring in 21-28% 50-59 year olds, 41-45% in 60-69 year olds, and 53-58% in patients over 70

79 (26). Dysplastic polyps which are left undetected can develop into colorectal cancer (CRC), the
80 third most prevalent cancer worldwide (27-30). There is interest, therefore, in understanding the
81 development of polyps as a precursor of cancer. Diverticular disease is the symptomatic
82 manifestation (normally abdominal pain) of people who have develop diverticula, which are
83 small bulges in the large intestine (31). Approximately 1 every 4 people with will develop
84 diverticulitis, which is the inflammation lead by bacteria and is associated with increased risk of
85 intestinal perforation (32).

86 **Material and Methods:**

87 **Study ethical approval and participants consent**

88 The ethics of this study were approved by the English National Health Service (NHS) Research
89 Ethics Committee in June 2015. Participants provided informed written consent after
90 registration and hold the right to drop out at any point of the study.

91 **Recruitment**

92 The TwinsUK ExHiBITT – Exploring Host microBIome inTeraction in Twins- cohort was
93 established between 2015 and 2018 to study interactions between colon microbiota and host
94 genomics. Twins were recruited from the TwinsUK cohort with the eligibility criteria outlined in
95 Supplementary_material_1.

96 Individuals who fell under this criterion were contacted by email. As the focus of our study was
97 healthy ageing, individuals were recruited from older age bands preferentially.

98 **Data and sample collection**

99 This cohort was annotated for three different types of phenotypes described in
100 Supplementary_material_2.

101 Living area was assigned by extracting Land Cover Map (LCM) 2015 1 km target class for each
102 of the participant's postcode using R package 'raster' and 'rgdal'. LCM classes were then
103 reassigned as urban, suburban or rural. Phenotypes were assessed thought self-reported
104 questionnaires in all cases except for weight and height in BMI, which were measured the day of

105 the visit. SocioEconomic Status (SES) was based on postcode location and assigned using
106 published deciles of the Index of Multiple Deprivation (IMD) for Scotland, Wales, England and
107 northern Ireland, where 1 is the most deprived and 10 is the least deprived (33). Frailty index was
108 annotated as described in Searle *et al.* (2008) (34).

109 Every patient underwent a colonoscopy, using the same bowel preparation (sennakot and
110 sodium picosulphate). Colon biopsies were taken at colonoscopies from up to four locations
111 (right colon, left colon, terminal ileum and cecum), caecal fluid, saliva and blood samples were
112 collected at time of visit (Supplementary_material_3). Stool samples were taken 24 hours prior,
113 before bowel preparation, and also at more than one week after the visit.

114 Data recorded just before commencing colonoscopy included presence/absence of irritable
115 bowel syndrome (IBS), and presence/absence of a history of abdominal pain, loose stool or
116 constipation. Phenotypic information collated during colonoscopy included endoscopic findings
117 (i.e. polyps and location and number of areas containing diverticulae), pain scores as assessed by
118 the endoscopist using the modified Gloucester scale (35) (1= comfortable, 5= frequent
119 discomfort with significant distress), quality of bowel preparation and time to caecum.
120 Histological outcomes from clinical biopsies of lesions were collated after the procedure.

121 **Immune profiling from peripheral blood and biopsies**

122 Peripheral blood mononuclear cells (PBMC) were isolated using ficoll-paque density gradient
123 centrifugation method. Multi-parametric flow cytometry was performed after staining with
124 relevant fluorescent monoclonal antibodies to quantify T cell. Effector memory T-cells were
125 identified as $CD3^+CD4^+CD25^-CD45RO^+CD45RA^-CCR7^-$, which then subsequently defined Th1
126 ($CXCR3^+CCR6^-$), Th17 ($CXCR3^-CCR6^+$), Th1-17 hybrid ($CXCR3^+CCR6^+$) and Th2 ($CXCR3^-$
127 $CCR6^-CCR4^+$) cells. Antigen experienced regulatory T cells (Ag Exp Treg) were defined as
128 $CD3^+CD4^+CD25^+CD45RA^-CCR4^+$ which were then subdivided into T helper like subsets based
129 on CCR6 and CXCR3 expression (Figure 1, panel a): 1A).

130 Endoscopically acquired colonic biopsies were sampled and partially disrupted by gently
131 compressing the epithelial/luminal aspect of the biopsy into the foam matrix. Complete culture
132 medium (supplemented with rhIL2, broad spectrum antibiotics and anti-fungal reagents) was
133 added and immune cells progressively migrated out of tissue into the culture medium. Cells were
134 harvested after 48 hours for downstream analysis. Leukocyte yield using this system was typically
135 in the region of 2×10^5 cells per biopsy. The cells were then stimulated with PMA and ionomycin
136 for 3 hours and analysed by intracellular cytokine staining and flow cytometry. T helper cell
137 subsets were defined as Th1 (IFN- γ^+ IL-17 $^-$), Th17 (IFN- γ IL-17 $^+$), and Th1-17 (IFN- γ^+ IL-17 $^+$)
138 cells. (Figure 1, panel b): 1B)

139 The data for each type of cell was calculated as a percentage of parent cell population and
140 analysed using graphpad prism software.

141 **Statistical analysis**

142 **Descriptive statistics** for sex, rearing, ethnicity, smoking status, living area and
143 socioeconomical status as well as polyp presence) and measured variables (BMI, age and frailty)
144 were calculated using RStudio (version 0.99.489 – © 2009-2015 RStudio, Inc).

145 For the concordance analysis of colonic traits, twin pairs where one of the individuals had
146 missing information were removed. The formula employed was: $CR_{pairwise} = (Number\ of\ concordant\ pairs / (Number\ of\ concordant\ pairs + number\ of\ discordant\ pairs)) * 100$ and $CR_{proband} = (2 * Number\ of\ concordant\ pairs / ((2 * Number\ of\ concordant\ pairs) + number\ of\ discordant\ pairs)) * 100$.

149 **Inferential statistics** were employed to interrogate the cohort through Linear Mixed Effect
150 Models (LMEM) using the algorithm provided in the R package lme4 (36). The model employed
151 was: $lmer(Trait \sim Frailty + Age + BMI + Quality_of_bowel_prep + (1 | Family_No))$, where the
152 random effects were the biological variates (frailty, age and BMI) and a technical covariate
153 (quality of bowel preparation). The fixed effect was family relatedness. The traits studied were
154 the four colonoscopy-derived phenotypes previously described. Moreover, a second model:
155 $lmer(Time_to_caecum \sim Pain_score + Endoscopist + AbdSym_including_IBS + Quality_of_bowel_prep$

156 + Age + Frailty + BMI + (1 | Family_No) was employed to identify any connexion between time
157 to caecum and the phenotypes measured. Bonferroni correction was applied to all the results
158 obtained from the statistical analysis.

159 Differences in *between* and *within* variation in twin pairs were studied using a linear model. The
160 model used was: $lmer(\text{Trait} \sim BMI^b + BMI^w + Frailty^b + Frailty^w + (1 | Family_No))$, where b
161 (*between*) denotes the mean for the trait in each family group, and w (*within*) the difference
162 between individuals and the family mean for each pair. Statistical difference in the *between* and
163 *within* coefficients for each trait was calculated using LINear COMBination of estimators
164 (LINCOM), implemented in STATA, where the model was reiterated.

165 **Data availability**

166 Data produced during the colonoscopy study will be publicly available through managed access.
167 Researchers interested can request access following TwinsUK procedure available at TwinsUK
168 Data Access Policy (<http://twinsuk.ac.uk>).

169 **Results and discussion:**

170 **1. Recruitment**

171 Two hundred and five twins volunteered for the study; out of those, two hundred successfully
172 completed the colonoscopy. Withdrawals were linked to the discovery of a suspected cancer
173 (n=3) or voluntary discontinuation during the intervention due to discomfort (n=2).

174 **2. Samples collection**

175 Colon biopsies were collected for interrogation of host genomics and microbiome analysis (data
176 not reported here). Samples were conserved in liquid nitrogen and included biopsies from i) left
177 colon (n=196), ii) terminal ileum (n=151), iii) caecum (n=73), and right colon (n=24) when one
178 of the other locations was difficult to sample. Mucosal biopsies to be used for microbial analysis
179 were conserved at -80°C. This included i) left colon (n=200), ii) right colon (n=179) and iii)
180 caecum (n=79). Colon biopsies were taken in triplicates. In total, 5 replicates of caecal fluid were
181 collected during the colonoscopy (n=197). Faecal samples were collected immediately prior to

182 bowel preparation (n=169), and one week after (n=188). Other samples included saliva (n=180)
183 and blood (n=204), which was stored as serum and plasma.

184 **3. Cohort descriptive statistical analysis**

185 The average age of the cohort was 58.70 ± 9.55 (F= 58.60 ± 9.52 , M= 59.04 ± 9.38), BMI was
186 26.37 ± 5.22 (F= 26.01 ± 5.18 , M= 27.66 ± 5.21), and frailty index 0.18 ± 0.10 (F= 0.19 ± 0.10 ,
187 M= 0.17 ± 0.09) (Supplementary_material_4, panel a). Twin pairs where differences between
188 continuous traits (i.e. BMI, frailty index and EIMD deciles was bigger than 1 standard deviation
189 were considered discordant (Supplementary_material_4, panel b). One twin pair was found
190 discordant for BMI, and 2 for frailty. In total, 39 twin pairs were found discordant for EIMD
191 decile. The twin pair with discordant BMI and frailty was also discordant for EIMD deciles.

192 In total, there were one hundred and sixty-one women and forty-four men in the cohort. Only
193 four individuals (2%) were reared apart. Ninety-five percent of the individuals identified
194 themselves as white, 2% as mixed, 2% as black and 1% as Asian. Five percent of twins could not
195 attend the colonoscopy visit with their co-twins, and one individual's twin dropped out from the
196 study just before the visit. Smokers represented 26% of the cohort, 66% of individuals never
197 smoked and 3% considered themselves as ex-smokers. Currently, 55% of the cohort live in the
198 same county as their co-twin. Fifty-nine percent of the cohort live in sub-urban areas, 28% in
199 rural areas and 11% in urban areas. Regarding socioeconomic status, 8% of the cohort were
200 classified as belonging to IMD decile 1-2, 15% to SES 3-5, 42% to SES 6-8 and 35% to SES 9-
201 10 where 10 is the least deprived (Supplementary_material_4, panel c).

202 In total, colonoscopy information from 196 individuals was collected. This information is next
203 described following the time sequence of the data collection and from specific to accumulative
204 phenotypic traits.

205 ***Pre-procedure outcomes***

206 Twenty individuals (13%) reported irritable bowel syndrome (IBS), with a concordance rate of
207 50% (proband) respectively. Everybody with IBS reported one type or another of abdominal

208 symptom. The different types of symptoms recorded were: i) pain/cramps (n=26), ii)
209 constipation (n=22), iii) rectal bleeding (n=2), diarrhoea (n=7) and alternative
210 diarrhoea/constipation (n=6). The accumulative trait '*presence of abdominal symptoms or IBS*'
211 counted for 40 individuals (28%) affected by at least one symptom. The pairwise concordance
212 rate was 48%, and the proband was 65% (Table 1). The influence of genetic and environmental
213 factors on the emergence of IBS has been the subject of considerable debate, with increasing
214 evidence that supports a role for genetic susceptibility. Our findings on concordance rates, is
215 slightly higher than other MZ twin studies which have showed concordance rates between 17%
216 and 33%(37-39).

217 ***Procedure related outcomes***

218 Out of the 196 individuals with colonoscopy information, 4 of them had poor bowel
219 preparation, 25 adequate and the rest had good bowel preparation. **Quality of bowel**
220 **preparation** was used in the LMEM as a potential confounder.

221 **Sedation** provided included midazolam, fentanyl, endotox or nothing. The medication index
222 was created considering the following factors: i) 1 mg of Midazolam= 1 index unit, ii) 25 µcg of
223 Fentanyl = 1 unit, and iii) Endotox use = 1 index unit. Sedation scores ranged from 1-6, average
224 3.8 ± 2.1 . 15 twin pairs were discordant by more than one SD, giving a proband concordance
225 rate equal to 91% (Table 1). These concordances should be taken with caution, as the
226 endoscopist was not blinded to twin pairing. Concordant twin pairs for sedation were selected to
227 study **pain scores** associations with time to caecum.

228 Two different types of pain traits were used, the original pain score taken during the
229 colonoscopy and the predicted pain score adjusted taking into account the sedation. For that
230 purpose, a Linear Mixed Effect Model ($Pain_score \sim Sedation_score + (1 | Family)$) was built in R to
231 calculate the residuals from pain score taking into account family and sedation. Concordance
232 rates were calculated in both traits, but only *Pain_score* of concordant twins for *Sedation_score* was
233 used in the model to calculate associations with time to caecum. The minimum pain score was 0,

234 and -1.19 in the predicted pain score. The maximum were 5 and 2.45 units respectively. The
235 average pain score was 1.58 ± 0.79 , and 0.003 ± 0.60 for the predicted pain score. The proband
236 concordant rate for pain score was 56% and 59% for the predicted pain score (Table 1).

237 Time to caecum was in average 12.97 ± 7.12 min, maximum time to caecum was 52 minutes and
238 minimum 1.33. There were 64 concordant pairs and 21 discordant by more than one standard
239 deviation. The concordance rate was 86% (proband)(Table 1). This minimal variation between
240 twins in caecal intubation time, suggests that technical difficulty and by inference colonic
241 morphology, was similar. Although this is not an entirely unexpected finding, it has not been
242 previously described in MZ twin colonoscopies. Focussing only on those individuals with
243 **polyps and/or diverticulosis**, in total 93 of them had one condition or both. Individuals had
244 between 0 to 4 total polyps and/or diverticulosis in total (Supplementary_material_5, panels b, f
245 and j). The proband concordance rate for polyps and/or diverticulosis was 74% (Table 1). This
246 concordance is illustrated in Supplementary_material_6.

247 Fifty-seven people had colonic **polyps** (28%), one of them had a potential cancer and
248 appropriate actions were taken. The number of polyps ranged from multiple (>7) to 1 (average
249 where present 1.5), (Supplementary_material_5, panels c, g and k). The concordance rate for
250 polyps was 42% (proband). Only 4 pairs were concordant for tubular adenomas, the rest of the pairs
251 discordant (n=29), giving a proband concordance rate of 22%.

252 Despite the fact that known **diverticular disease** was an exclusion for the study (due to the
253 increased risk of bowel perforation), 51 people were found to have diverticulosis on endoscopy
254 (26%), of which the majority (29) where located in the left colon. The number of locations for
255 diverticulae within an individual ranged from 0 to 2. No individuals had evidence of inflamed
256 diverticulae (diverticulitis). Twenty-one twin pairs (n=42) were concordant for diverticulosis and
257 nine cases (n=9) were discordant (Fig 3, panels: I, j, k, and i). Thus, *diverticulosis* had the highest
258 concordance between twins at 82% (proband) (Table 1, Supplementary_material_6).

259 In a previous twin cohort study from the Swedish Twin Registry, the MZ concordant rate for
260 diverticulosis was 6% only, due to the fact there were over fourteen fold times more discordant
261 twins for diverticulosis than concordant ones (40). Similarly, the diverticulosis study from the
262 Danish Twin Registry found that the diverticulosis twin concordance rate was 8% (40, 41).
263 Differences between these studies and the results from the colonoscopy TwinsUK is most likely
264 to be a function of ascertainment. Our participants were selected not to have a known diagnosis
265 of diverticulosis, and presence was ascertained endoscopically. Whereas these other studies relied
266 on health record data from physician diagnosis and asymptomatic co-twins may not have
267 undergone a colonoscopy. Alternatively there could come from environmental and genetic
268 variation between Scandinavian and British populations, or differences in advances in the
269 colonoscopy techniques (where employed), cohort size and recruitment criteria and timing of the
270 study (the Swedish Twin Registry took data from 1886 to 1980, and the Danish went from 1977
271 to 2011, while the TwinsUK colonoscopy study examined volunteers between 2015 and 2018).
272 Heritability of diverticular disease has been estimated by Strate and colleagues (2013) (42) as
273 53%, which could be an underestimate due to asymptomatic disease. To the best of our
274 knowledge, the high endoscopic concordance rate for diverticulosis in identical twins identified
275 in this cohort was never reported before. This indicates that genetic variants could contribute to
276 the development of diverticulosis, as previously indicated.

277 **Complications**

278 Despite the large number of samples collected, there were no major complications, including
279 perforation or bleeding. Minor complications included incomplete procedures secondary to
280 patient discomfort (n=2) or presence of a fixed sigmoid that limited endoscopic progression.
281 One 61year-old patient who received sedation experienced a transient vasovagal episode during
282 the procedure.

283 ***Post-procedure related outcomes***

284 One hundred and four individuals had some sort of **abnormality** in the mucosa observed either
285 during the colonoscopy or at histology. Individuals with presence of any abnormality represented
286 51% of the cohort, 45% of the cohort was absent of any sort of abnormality and the remaining
287 are those individuals with no colonoscopy information available (N/A). Abnormalities ranged
288 from 0 to 4 (Supplementary_material_5, panels a, e and i). Twenty-one twin pairs (n=42) were
289 discordant for any abnormality, and 41 twin pairs were concordant. This gave a concordant rate
290 of 80% (proband).

291 **4. Cohort inferential statistical analysis**

292 A LMEM was used to interrogate if our colonoscopy traits were associated with biological
293 covariates (i.e. BMI, age, frailty). Results from the LMEM showed that age and BMI were
294 statistically significant according to i) *total number of abnormalities*, ii) *total number of polyps and*
295 *diverticulosis*, and iii) *total number of polyps*. Total number of diverticulosis was only relatively closed
296 to be significant in the case of BMI (Supplementary_material_7, Table 2). There was no
297 detectable association between time to caecum and biological covariates.

298 Furthermore, between family (*b*) and within pair (*w*) twin differences for BMI and frailty index
299 were studied using linear models. No significance was found for frailty. Reflecting the results
300 above, and consistent with previous published studies (43-46), BMI_{*b*} was statistically significant
301 in all the traits studied such that higher BMI led to greater risk of anomalies. BMI difference
302 within pairs was significantly different from the between family difference in all four tests and
303 showed significant opposite relationship in the traits i) *total number of polyps*, and ii) *total number of*
304 *polyps and/or diverticulosis*, i.e. higher BMI within pairs led to reduced risk of anomalies. This could
305 indicate common factors to both twins, such as genetics and early life environment, could be the
306 link between with BMI and the colonoscopy traits studied such as polyps (Table 3), rather than
307 BMI being directly causal. This is intriguing given the evidence of host genetic factors impacting
308 the gut microbiome (47), and obesity (48). Only a minority of studies have looked at microbiome
309 as a potential biomarker associated with the development of polyps in healthy individuals (30, 46,

310 49). Further work with ExHiBITT will consider microbiome composition in relationship to
311 polyps and diverticular disease.

312 **Immune profiling outcomes**

313 The twin pairs were highly concordant for different T cell subsets in both blood (Figure 2, panel
314 a) and gut (Figure 3, panel a). Preliminary analysis showed differences in the immune response
315 between males and females such as increased CD4 proportion and reduced antigen experienced
316 Treg in females (not shown). Interestingly we found increased proportion of Th17 and Th2 cells
317 in the peripheral blood in autumn-winter seasons compared to spring- summer seasons (Figure
318 2, panel b). No marked differences were seen in the peripheral blood immune profile in traits
319 such as polyp or diverticulosis. However, increased proportion of hybrid Th1-17 cells producing
320 both IFN gamma and IL-17 were found in colonic biopsies from patients with diverticulosis
321 (Figure 3, panel b). No differences were found in the gut immune profile of individuals
322 with/without polyps. Since generation of effector T-cell responses has been shown to be
323 dependent on the composition of the intestinal microbiota, it will be interesting to look at the
324 microbiome driving these differences in our cohort.

325 **Conclusions:**

326 This cohort represents a great potential to study microbiome-host interactions in the colon, and
327 their implications for the host immune system. The cohort is annotated for a large number of
328 phenotypes representative of UK society. Preliminary findings showed that polyps are strongly
329 correlated with BMI and age, but that the relationship with BMI may be confounded by factors
330 genetics and other factors shared by twins. There is a high rate of concordance between twin
331 pairs for diverticulosis, less so for polyps. Interestingly, similar intubation times and pain scores
332 were found for twin pairs, which could indicate that familial factors determine the ease of
333 colonoscopy for both the endoscopist and patient. Further studies will include the high

334 throughput analysis of the samples. We have also successfully phenotyped immune profile from
335 the blood and gut of healthy twin pairs. High rate of concordance was found among twin pairs
336 for effector and regulatory T cell subsets highlighting genetic control of immune response in
337 monozygotic twins whereas seasonal variations found in the proportion of effector cell subsets
338 ascertains the environmental programming of immune responses. Hybrid Th1-17 cells in the gut
339 were shown to be associated with diverticulosis. Further analysis of this cohort will reveal the
340 ileal microbiota responsible for driving systemic and mucosal immune response.

341 **Abbreviations**

342 BMI, Body Mass Index

343 MZ, Monozygotic

344 NGS, Next Generation Techniques

345 NHS, National Health Service

346 SES socioeconomic status

347 **Declarations**

348 The authors declare they do not have conflict of interest.

349 **Acknowledgments**

350 The authors thank the twins for their participation in the study, and the Medical Research
351 Council (MRC) for funding this research (RE10740). We also wish to thank Clare Stockwell,
352 Rachel Horsfall and Isabelle Granville Smith from the Microbiome Project, and Genevieve
353 Lachance, Darioush Yarand and Merve Demiroglu from IT/Data & Administration resources,
354 King's College, University of London, for their technical assistance. Finally, we would also like to
355 thank to Dr Julia El-Sayed Moustafa for the advice provided with the statistical models.

356 **Funding**

357 This work was supported by a Medical Research Council (MRC) grant [grant number RE10740].

358 The TwinsUK study was funded by the Wellcome Trust and European Community's Seventh
359 Framework Programme (FP7/2007-2013). The TwinsUK study also receives support from the

360 National Institute for Health Research (NIHR)- funded BioResource, Clinical Research Facility
 361 and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust in
 362 partnership with King's College London.

363 **Authors' contributions**

364 MMO wrote the first draft of the manuscript, compiled the metadata and created the figures. HI
 365 contributed to collect colonoscopy data and contribute to the gastroenterological aspects of the
 366 manuscript. SHK carried out all the immunological analysis and contribute to write the
 367 manuscript. RB compiled the metadata related to socioeconomic status, frailty and geographical
 368 location. NP conducted the colonoscopies. TS, KS and CS conceived the idea and supervised
 369 the work. All authors contributed to the experimental plan, supervised the work and contributed
 370 to write the manuscript. All authors have approved the final manuscript.

371 **List of tables:**

372 **Table 1.** Concordance rate expressed in percentage.

Trait	C	D	T	Concordant rate % (pairwise) (C/(C+D))*100	Concordant rate (proband) (2C/(2C+D))*100
Any abnormality in mucosa	41	21	62	66%	80%
P and/or Di	34	24	58	59%	74%
Polyps	12	33	45	27%	42%
Tubular adenoma	4	29	33	12%	22%
Diverticulosis	21	9	30	70%	82%
Abdominal symp	13	14	27	48%	65%
IBS	5	10	15	33%	50%
Time to caecum	64	21	85	65%	86%
Pain score	16	25	41	39%	56%
Predicted pain	17	24	41	41%	59%
Medication	76	15	91	84%	91%

373 C= Number of concordant pairs, D=Number of Discordant pairs, T= Total number of pairs,

374 P=polyps, Di=diverticulosis.

375

376 **Table 2:** Results from the LMEM used to interrogate the phenotypes from the colonoscopy
 377 analysis according to the biological covariates: BMI, age and frailty:

378

Number of Abnormalities (n=190)	Estimate	Std.	t value	Pr(>Chisq)
---------------------------------	----------	------	---------	------------

		Error		
Frailty Index	-1.09	0.68	-1.61	0.10
Age	0.03	0.01	3.60	***
BMI	0.04	0.01	3.15	**
Number of Polyps and Diverticulosis (n=190)				
Frailty Index	-0.97	0.61	-1.60	0.11
Age	0.03	0.01	4.04	***
BMI	0.05	0.01	4.04	***
Number of Polyps (n=188)				
Frailty I	-0.81	0.52	-1.56	1
Age	0.02	0.01	3.30	**
BMI	0.03	0.01	3.16	**
Number of Diverticulosis locations (n=183)				
Frailty I	-0.37	0.28	-1.33	0.19
Age	0.01	0.01	2.02	*
BMI	0.01	0.01	1.52	0.13

379 Signif. codes: <0.001 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘.’ 1; **Bonferroni correction: 0.0125**

380

381 **Table 3:** Results from the LMEM and LINCOM test used to interrogate the between and within

382 variation in BMI and frailty:

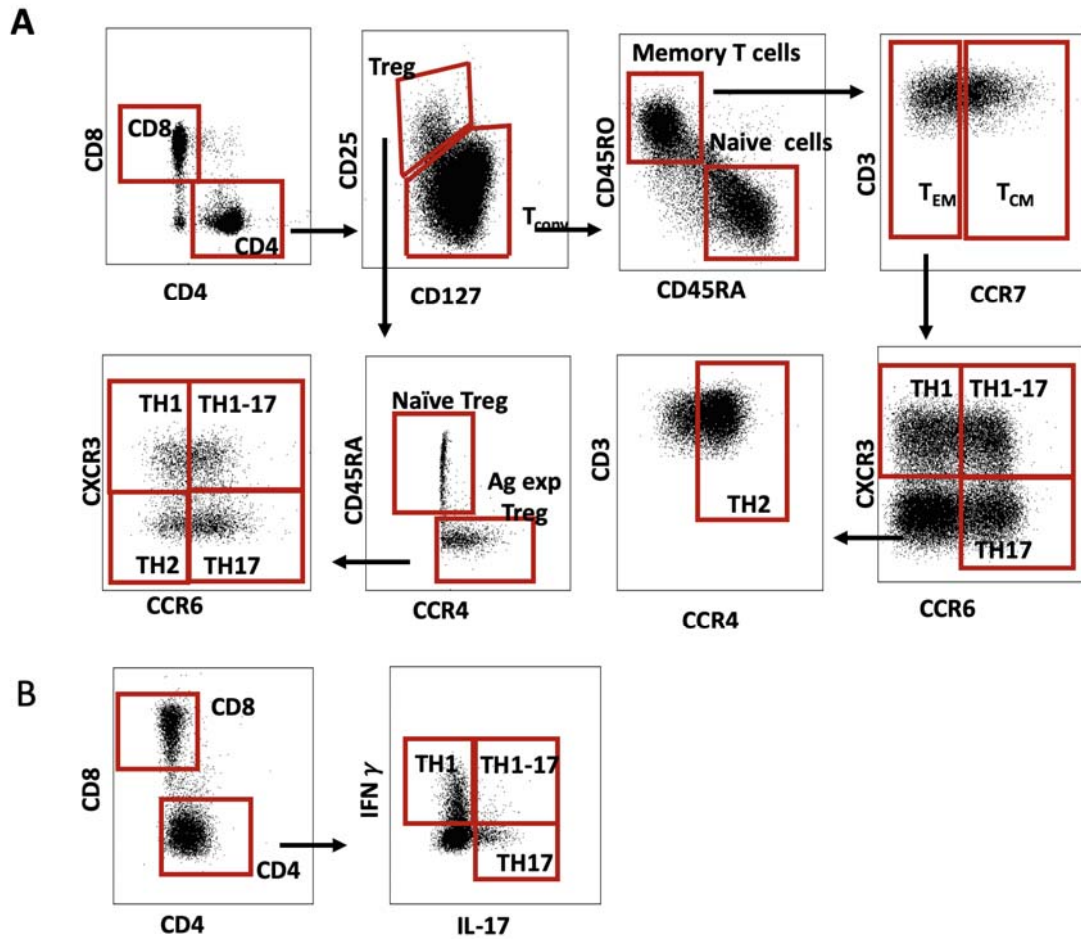
383

	Estimate	Std. Error	t value	Pr(>Chisq)	LINCOM
Number of Abnormalities (n=184)					
BMI ^b	0.03	0.01	3.50	0.001 **	0.005 * n/a
BMI ^w	-0.00	0.01	-0.16	0.876	
Frailty ^b	-0.23	0.52	-0.43	0.668	
Frailty ^w	0.16	0.50	0.32	0.750	
Number of Polyps and Diverticulosis (n=184)					
BMI ^b	0.05	0.01	3.42	0.001 **	0.001 ** n/a
BMI ^w	-0.06	0.03	-2.37	0.020 .	
Frailty ^b	-0.86	0.73	-1.18	0.242	
Frailty ^w	0.88	0.83	1.07	0.288	
Number of Polyps (n=182)					
BMI ^b	0.03	0.01	2.23	0.028 .	0.003 * n/a
BMI ^w	-0.06	0.02	-2.38	0.019 .	
Frailty ^b	-0.82	0.65	-1.26	0.211	
Frailty ^w	0.11	0.88	0.12	0.904	
Number of Diverticulosis (n=178)					
BMI ^b	0.02	0.01	2.25	0.027 .	0.05 . n/a
BMI ^w	-0.01	0.01	-0.79	0.430	
Frailty ^b	0.28	0.67	0.42	0.678	
Frailty ^w	0.71	0.38	1.88	0.063	

384 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1; **Bonferroni correction: 0.0125,**

385 LINCOM: LINear COMbinations of estimators

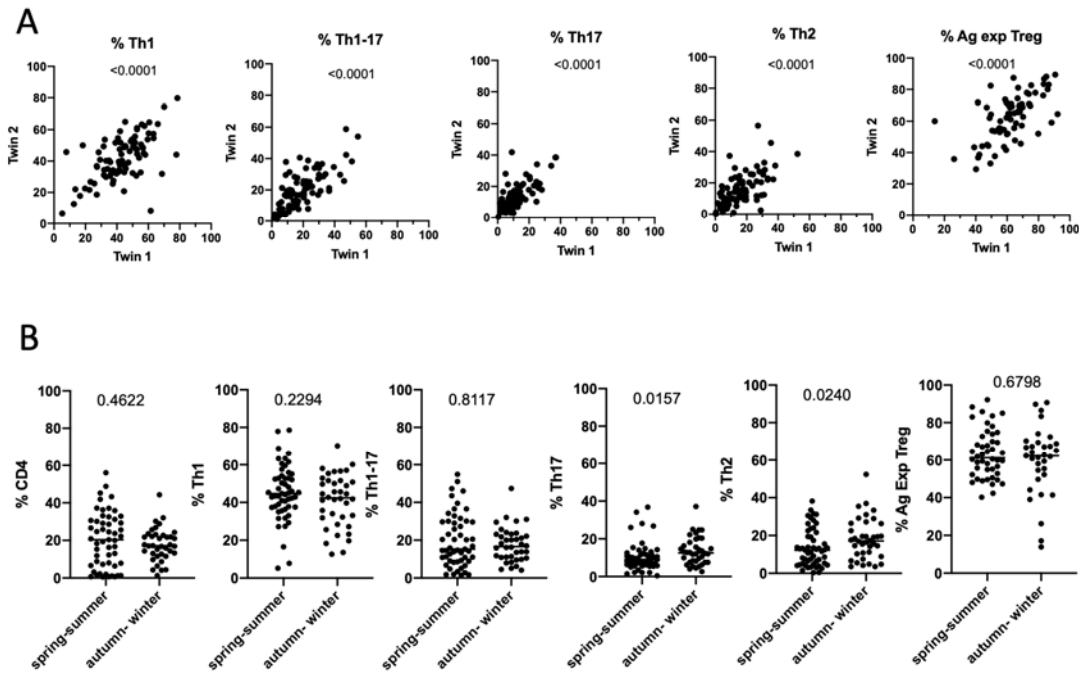
386 **List of figures:**



387

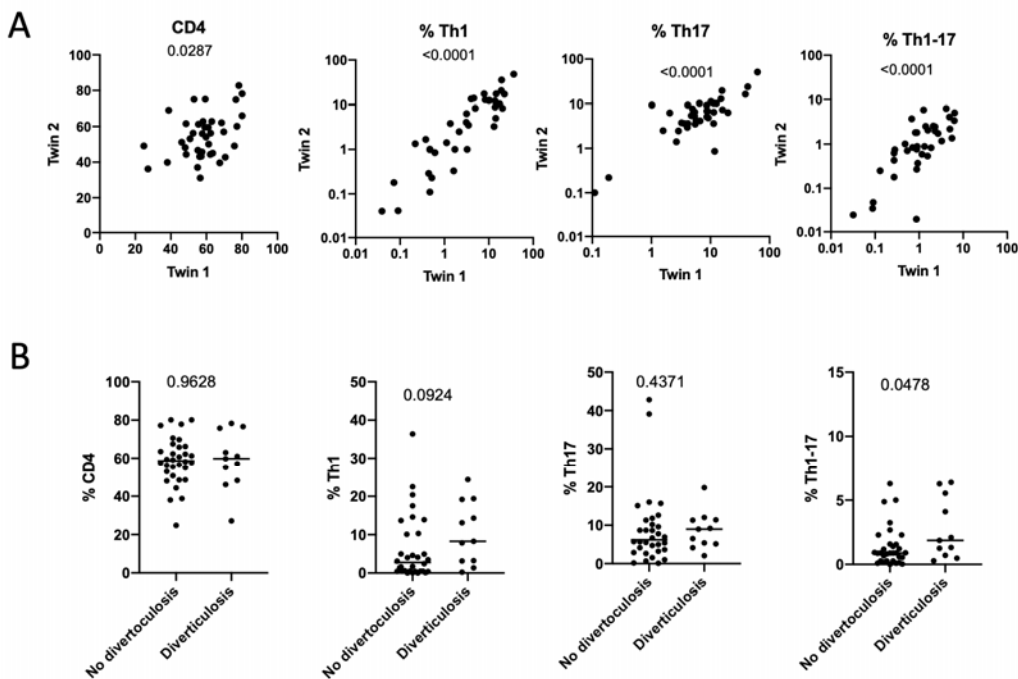
388 **Figure 1. Flow cytometric gating strategy.** Panel a) Flowcytometric analysis of peripheral
 389 blood CD4 T cells (gated on CD3⁺ live lymphocytes) which were then divided into
 390 CD127^{high}CD25^{low} conventional T cells (T_{conv}) and CD127^{low}CD25^{high} regulatory T cells (Treg).
 391 T_{conv} cells were then divided into naive and memory T cells. CD45RO⁺CD45RA⁻ memory T
 392 cells were subdivided into CCR7⁻ effector memory (TEM) and CCR7⁺ central memory (TCM) T
 393 cells. TEM defined Th17 (CCR6⁺CXCR3⁻), Th1 (CXCR3⁺CCR6⁺), Th1-17 (CXCR3⁺CCR6⁺)
 394 and Th2 (CXCR3⁻CCR6⁻CCR4⁺) cells. Antigen experienced Treg (Ag exp Treg) were defined as
 395 CD45RA⁻CCR4⁺ Treg which were then subdivided into T helper like subsets based on CCR6
 396 and CXCR3 expression. Panel b) Flow cytometric analysis of lamina propria mononuclear cells-
 397 CD4 T cells (gated on CD45⁺CD3⁺ live lymphocytes) were divided into Th1, Th1-17 and Th17
 398 cells based on IFN gamma and Il-17 expression.

399



400

401 **Figure 2. Peripheral blood immunophenotyping.** Panel a) Proportion of different T helper
402 cell subsets correlate between individual Twin pairs. Panel b) Frequency of CD4 T cells, Th1,
403 Th1-17, Th17, Th2 and Ag exp Tregs between samples collected at different seasons.



404

405 **Figure 3. Gut immunophenotyping.** Panel a) Proportion of different T helper cell subsets in
406 the gut correlate between Twin pairs. Panel b) Proportion of CD4 T cells, Th1, Th1-17 and
407 Th17 between individuals with or without diverticulosis.

408 **Bibliography**

- 409 1. Agur A, Lee M, Grant J. Grant's Atlas of Anatomy. 10th Ed. . London, UK:
410 Lippincott Williams and Wilkins; 1999.
- 411 2. McGregor A, Decker G. Lee McGregor's synopsis of surgical anatomy. 12th ed.
412 Bristol: John Wright; 1986.
- 413 3. Krogh D. Biology: A Guide to the Natural World: Benjamin-Cummings Publishing
414 Company; 2010.
- 415 4. Burrows MP, Volchkov P, Kobayashi KS, Chervonsky AV. Microbiota regulates type
416 1 diabetes through Toll-like receptors. 2015;112(32):9973-7.
- 417 5. Suau A, Bonnet R, Sutren M, Godon J-J, Gibson GR, Collins MD, et al. Direct
418 Analysis of Genes Encoding 16S rRNA from Complex Communities Reveals Many Novel
419 Molecular Species within the Human Gut. 1999;65(11):4799-807.
- 420 6. Savage DC. Microbial Ecology of the Gastrointestinal Tract. 1977;31(1):107-33.
- 421 7. Andersson AF, Lindberg M, Jakobsson H, Bäckhed F, Nyrén P, Engstrand L.
422 Comparative Analysis of Human Gut Microbiota by Barcoded Pyrosequencing. PLOS ONE.
423 2008;3(7):e2836.
- 424 8. Yatsunenkov T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, et
425 al. Human gut microbiome viewed across age and geography. Nature. 2012;486:222.
- 426 9. The Human Microbiome Project C, Huttenhower C, Gevers D, Knight R, Abubucker
427 S, Badger JH, et al. Structure, function and diversity of the healthy human microbiome.
428 Nature. 2012;486:207.
- 429 10. Bloom SM, Bijanki VN, Nava GM, Sun L, Malvin NP, Donermeyer DL, et al.
430 Commensal Bacteroides species induce colitis in host-genotype-specific fashion in a mouse
431 model of inflammatory bowel disease. Cell host & microbe. 2011;9(5):390-403.
- 432 11. Bottacini F, Ventura M, van Sinderen D, O'Connell Motherway M. Diversity, ecology
433 and intestinal function of bifidobacteria. Microbial cell factories. 2014;13 Suppl 1(Suppl
434 1):S4-S.
- 435 12. Ruff WE, Kriegel MA. Autoimmune host–microbiota interactions at barrier
436 sites and beyond. Trends in Molecular Medicine. 2015;21(4):233-44.
- 437 13. Belkaid Y, Naik S. Compartmentalized and systemic control of tissue immunity by
438 commensals. Nature Immunology. 2013;14:646.
- 439 14. Steegenga WT, de Wit NJ, Boekschoten MV, Ijssennagger N, Lute C, Keshtkar S, et
440 al. Structural, functional and molecular analysis of the effects of aging in the small intestine
441 and colon of C57BL/6J mice. BMC Med Genomics. 2012;5:38-.
- 442 15. Moayyeri A, Hammond CJ, Hart DJ, Spector TD. The UK Adult Twin Registry
443 (TwinsUK Resource). Twin research and human genetics : the official journal of the
444 International Society for Twin Studies. 2013;16(1):144-9.
- 445 16. Verdi S, Abbasian G, Bowyer RCE, Lachance G, Yarand D, Christofidou P, et al.
446 TwinsUK: The UK Adult Twin Registry. Twin research and human genetics : the official
447 journal of the International Society for Twin Studies. 219.
- 448 17. Goodrich Julia K, Waters Jillian L, Poole Angela C, Sutter Jessica L, Koren O,
449 Blehman R, et al. Human Genetics Shape the Gut Microbiome. Cell. 2014;159(4):789-99.

- 450 18. Grundberg E, Small KS, Hedman ÅK, Nica AC, Buil A, Keildson S, et al. Mapping
451 cis- and trans-regulatory effects across multiple tissues in twins. *Nature genetics*.
452 2012;44(10):1084-9.
- 453 19. Jackson MA, Jeffery IB, Beaumont M, Bell JT, Clark AG, Ley RE, et al. Signatures
454 of early frailty in the gut microbiota. *Genome Medicine*. 2016;8(1):8.
- 455 20. Jackson MA, Verdi S, Maxan M-E, Shin CM, Zierer J, Bowyer RCE, et al. Gut
456 microbiota associations with common diseases and prescription medications in a population-
457 based cohort. *Nature Communications*. 2018;9(1):2655.
- 458 21. Carlin JB, Sterne JAC, Gurrin LC, Morley R, Dwyer T. Regression models for twin
459 studies: a critical review. *International Journal of Epidemiology*. 2005;34(5):1089-99.
- 460 22. Ding X, Hu Y, Guo X, Guo X, Morgan I, He M. Possible Causes of Discordance in
461 Refraction in Monozygotic Twins: Nearwork, Time Outdoors and Stochastic Variation.
462 *Invest Ophthalmol Vis Sci*. 2018;59(13):5.
- 463 23. Boomsma D, Busjahn A, Peltonen L. Classical twin studies and beyond. *Nature*
464 *Reviews Genetics*. 2002;3:10.
- 465 24. Yao MD, von Rosenvinge EC, Groden C, Mannon PJ. Multiple endoscopic biopsies
466 in research subjects: safety results from a National Institutes of Health series. *Gastrointestinal*
467 *endoscopy*. 2009;69(4):906-10.
- 468 25. Lieberman DA, Faigel DO, Logan JR, Mattek N, Holub J, Eisen G, et al. Assessment
469 of the quality of colonoscopy reports: results from a multicenter consortium. *Gastrointestinal*
470 *endoscopy*. 2009;69(3):645-53.
- 471 26. Pezzoli A, Matarese V, Rubini M, Simoni M, Caravelli G, Stockbrugger R, et al.
472 Colorectal cancer screening: results of a 5-year program in asymptomatic subjects at
473 increased risk. *Digestive and liver disease*. 2007;39(1):33-9.
- 474 27. Sinicrope PS, Goode EL, Limburg PJ, Vernon SW, Wick JB, Patten CA, et al. A
475 Population-Based Study of Prevalence and Adherence Trends in Average Risk Colorectal
476 Cancer Screening, 1997 to 2008. 2012;21(2):347-50.
- 477 28. FriedenberG FK, Singh M, George NS, Sankineni A, Shah S. Prevalence and
478 distribution of adenomas in black Americans undergoing colorectal cancer screening.
479 *Digestive diseases and sciences*. 2012;57(2):489-95.
- 480 29. Blumenstein I, Tacke W, Bock H, Filmann N, Lieber E, Zeuzem S, et al. Prevalence
481 of colorectal cancer and its precursor lesions in symptomatic and asymptomatic patients
482 undergoing total colonoscopy: results of a large prospective, multicenter, controlled
483 endoscopy study. 2013;25(5):556-61.
- 484 30. Brim H, Yooshep S, Zoetendal EG, Lee E, Torralbo M, Laiyemo AO, et al.
485 Microbiome analysis of stool samples from African Americans with colon polyps. *PLoS one*.
486 2013;8(12):e81352-e.
- 487 31. Pemberton JH. Clinical manifestations and diagnosis of acute diverticulitis in adults.
488 In: Lamont JT, Grover S, editors. *uptodate2019*. p. <https://www.uptodate.com>.
- 489 32. Böhm SK. Risk Factors for Diverticulosis, Diverticulitis, Diverticular Perforation, and
490 Bleeding: A Plea for More Subtle History Taking. *Viszeralmedizin*. 2015;31(2):84-94.
- 491 33. Bowyer RCE, Jackson MA, Le Roy CI, Ni Lochlainn M, Spector TD, Dowd JB, et al.
492 Socioeconomic Status and the Gut Microbiome: A TwinsUK Cohort Study. *Microorganisms*.
493 2019;7:17.
- 494 34. Searle SD, Mitnitski A, Gahbauer EA, Gill TM, Rockwood KJBG. A standard
495 procedure for creating a frailty index. 2008;8(1):24.
- 496 35. Chilton A, Rutter M. Quality assurance guidelines for colonoscopy. *NHS Cancer*
497 *Screening Programmes*. 2011.
- 498 36. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using
499 *lme4*. 2015. 2015;67(1):48 %J *Journal of Statistical Software*.

- 500 37. Morris-Yates A, Talley NJ, Boyce PM, Nandurkar S, Andrews G. Evidence of a
501 genetic contribution to functional bowel disorder. *The American journal of gastroenterology*.
502 1998;93(8):1311-7.
- 503 38. Levy RL, Jones KR, Whitehead WE, Feld SI, Talley NJ, Corey LA. Irritable bowel
504 syndrome in twins: heredity and social learning both contribute to etiology. *Gastroenterology*.
505 2001;121(4):799-804.
- 506 39. Mohammed I, Cherkas LF, Riley SA, Spector TD, Trudgill NJ. Genetic influences in
507 irritable bowel syndrome: a twin study. *The American journal of gastroenterology*.
508 2005;100(6):1340.
- 509 40. Granlund J, Svensson T, Olén O, Hjern F, Pedersen NL, Magnusson PKE, et al. The
510 genetic influence on diverticular disease – a twin study. *Alimentary Pharmacology &
511 Therapeutics*. 2012;35(9):1103-7.
- 512 41. Strate LL, Erichsen R, Baron JA, Mortensen J, Pedersen JK, Riis AH, et al.
513 Heritability and Familial Aggregation of Diverticular Disease: A Population-Based Study of
514 Twins and Siblings. *Gastroenterology*. 2013;144(4):736-42.e1.
- 515 42. Reichert MC, Lammert F. The genetic epidemiology of diverticulosis and diverticular
516 disease: Emerging evidence. *United European gastroenterology journal*. 2015;3(5):409-18.
- 517 43. Comstock SS, Hortos K, Kovan B, McCaskey S, Pathak DR, Fenton JI. Adipokines
518 and obesity are associated with colorectal polyps in adult males: a cross-sectional study. *PloS
519 one*. 2014;9(1):e85939-e.
- 520 44. Ben Q, An W, Jiang Y, Zhan X, Du Y, Cai QC, et al. Body Mass Index Increases
521 Risk for Colorectal Adenomas Based on Meta-analysis. *Gastroenterology*. 2012;142(4):762-
522 72.
- 523 45. Omata F, Deshpande GA, Ohde S, Mine T, Fukui T. The association between obesity
524 and colorectal adenoma: systematic review and meta-analysis. *Scandinavian Journal of
525 Gastroenterology*. 2013;48(2):136-46.
- 526 46. Rex DK, Lehman GA, Ulbright TM, Smith JJ, Pound DC, Hawes RH, et al. Colonic
527 neoplasia in asymptomatic persons with negative fecal occult blood tests: influence of age,
528 gender, and family history. *The American Journal of Gastroenterology*. 1993;88(6):6.
- 529 47. Goodrich JK, Davenport ER, Beaumont M, Jackson MA, Knight R, Ober C, et al.
530 Genetic Determinants of the Gut Microbiome in UK Twins. *Cell host & microbe*.
531 2016;19(5):731-43.
- 532 48. Bäckhed F, Manchester JK, Semenkovich CF, Gordon JI. Mechanisms underlying the
533 resistance to diet-induced obesity in germ-free mice. *Proceedings of the National Academy of
534 Sciences of the United States of America*. 2007;104(3):979-84.
- 535 49. Hale VL, Chen J, Johnson S, Harrington SC, Yab TC, Smyrk TC, et al. Shifts in the
536 Fecal Microbiota Associated with Adenomatous Polyps. *Cancer epidemiology, biomarkers &
537 prevention* : a publication of the American Association for Cancer Research, cosponsored by
538 the American Society of Preventive Oncology. 2017;26(1):85-94.
539