

1 **First Insights into microbial changes within an Inflammatory Bowel Disease Family**  
2 **Cohort study**

3 Philipp Rausch<sup>1\*</sup>, Ilka Ratjen<sup>2,3,4\*</sup>, Lukas Tittmann<sup>2,3</sup>, Janna Enderle<sup>2,5</sup>, Eike Matthias  
4 Wacker<sup>1</sup>, Kathrin Jaeger<sup>2,3</sup>, Malte Christoph Rühlemann<sup>1</sup>, Katrin Franzpötter<sup>2</sup>, Pierre Ellul<sup>6</sup>,  
5 Robert Kruse<sup>7</sup>, Jonas Halfvarson<sup>7</sup>, Dirk Roggenbuck<sup>8,9</sup>, David Ellinghaus<sup>1</sup>, Gunnar Jacobs<sup>2,3</sup>,  
6 Michael Krawczak<sup>10</sup>, Stefan Schreiber<sup>1,11</sup>, Corinna Bang<sup>1</sup>, Wolfgang Lieb<sup>2,3\*</sup>, Andre Franke<sup>1\*</sup>

7 \*denotes equal contribution

8 <sup>1</sup>Institute of Clinical Molecular Biology, University Hospital Schleswig-Holstein, University  
9 of Kiel, Kiel, Germany

10 <sup>2</sup>Institute of Epidemiology, University Hospital Schleswig-Holstein, University of Kiel, Kiel,  
11 Germany

12 <sup>3</sup>popgen biobank, University Hospital Schleswig-Holstein, University of Kiel, Kiel, Germany

13 <sup>4</sup>Department of Internal Medicine II, University Cancer Center Schleswig-Holstein,  
14 University Hospital Schleswig-Holstein, Kiel, Germany

15 <sup>5</sup>Institute of Human Nutrition and Food Science, University of Kiel, Kiel, Germany

16 <sup>6</sup>Division of Gastroenterology, Department of Medicine, Mater Dei Hospital, Malta

17 <sup>7</sup>Department of Gastroenterology, Faculty of Medicine and Health, Örebro University,  
18 Örebro, Sweden

19 <sup>8</sup>Institute of Biotechnology, Faculty Environment and Natural Sciences, Brandenburg  
20 University of Technology Cottbus-Senftenberg, Germany

21 <sup>9</sup>Medipan/GA Generic Assays GmbH, Ludwig-Erhard-Ring 3, 15827 Dahlewitz, Germany

22 <sup>10</sup>Institute of Medical Informatics and Statistics, University Hospital Schleswig-Holstein,  
23 University of Kiel, Kiel, Germany

24 <sup>11</sup>Department of Internal Medicine I, University Hospital Schleswig-Holstein, Kiel, Germany

26 **Correspondence: Prof. Dr. Andre Franke**, Christian-Albrechts-University of Kiel,  
27 University Hospital Schleswig-Holstein, Institute of Clinical Molecular Biology, Rosalind-  
28 Franklin-Str. 12, 24105 Kiel, Germany; [a.franke@ikmb.uni-kiel.de](mailto:a.franke@ikmb.uni-kiel.de)

29 **Abbreviations used:** CD-Crohn's disease; IBD-inflammatory bowel disease; IC-informed  
30 consent; UC-ulcerative colitis; uIBD-undetermined inflammatory bowel disease, MD-index-  
31 Microbiome dysbiosis index; GMHI-General Microbiome Health Index; PRS-Polygenic Risk

32 **NOTE:** This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.  
Scores

33 **ABSTRACT**

34 **Background:** The prospective Kiel Inflammatory Bowel Disease (IBD) Family Cohort Study  
35 (KINDRED cohort) was initiated in 2013 to systematically and extensively collect data and  
36 biosamples from index IBD patients and their relatives (e.g., blood, stool), a population at  
37 high risk for IBD development. Regular follow-ups were conducted to collect updated health  
38 and lifestyle information, to obtain new biosamples, and to capture the incidence of IBD  
39 during development. By combining taxonomic and imputed functional microbial data  
40 collected at successive time points with extensive anthropometric, medical, nutritional, and  
41 social information, this study aimed to characterize the factors influencing the microbiota in  
42 health and disease via detailed ecological analyses.

43 **Results:** Using two dysbiosis metrics (MD-index, GMHI) trained on the German KINDRED  
44 cohort, we identified strong and generalizable gradients within and across different IBD  
45 cohorts, which correspond strongly with IBD pathologies, physiological manifestations of  
46 inflammation (e.g., Bristol stool score, ASCA IgA/IgG), genetic risk for IBD, and general risk  
47 of disease onset. Anthropometric and medical factors influencing transit time strongly modify  
48 bacterial communities. Various *Enterobacteriaceae* (e.g., *Klebsiella sp.*) and opportunistic  
49 *Clostridia* pathogens (e.g., *C. XIVa clostridioforme*), characterize in combination with ectopic  
50 oral taxa (e.g. *Veillonella sp.*, *Cand. Saccharibacteria sp.*, *Fusobacterium nucleatum*) the  
51 distinct and chaotic IBD-specific communities. Functionally, amino acid metabolism and  
52 flagellar assembly are beneficial, while mucolytic functions are associated with IBD.

53 **Conclusions:** Our findings demonstrate broad-scale ecological patterns which indicate drastic  
54 state transitions of communities into characteristically chaotic communities in IBD patients.  
55 These patterns appear to be universal across cohorts and influence physiological signs of  
56 inflammation, display high resilience, but show only little heritability/intrafamily  
57 transmission.

58

59 **Keywords:** Kiel Inflammatory Bowel Disease Family cohort, epidemiology, microbiota, 16S,  
60 dysbiosis

61

## 62 **INTRODUCTION**

63 Inflammatory bowel disease (IBD), with the most common manifestations of Crohn’s disease  
64 (CD) and ulcerative colitis (UC), is characterized by chronic, relapsing inflammation of the  
65 gastrointestinal tract [1] resulting from a complex interplay of genetic, lifestyle, and  
66 environmental factors [2–4]. The incidence and prevalence of IBD are rising in many parts of  
67 the world [5,6], and the highest incidences are currently reported in North America and  
68 Europe [6,7], while short-term projections expect a prevalence of 1% in high-income  
69 countries by the year 2030 [8].

70 Clustering of IBD cases within families is not uncommon, suggesting a significant  
71 contribution of shared genes and lifestyle or environmental factors to IBD susceptibility [4,9–  
72 12]. First-degree relatives of IBD patients have a 3- to 20-fold greater risk of developing the  
73 disease themselves than does the general population [4,11,13,14]. In particular, for siblings of  
74 IBD patients, estimates of the relative disease risk range from 10 to 50 [15]. Meta-analyses  
75 and genome-wide association studies (GWAS) have identified 320 loci associated with either  
76 CD or UC or both [16]. Most disease-related genes and pathways are thought to influence  
77 disease risk by altering the epithelial barrier, the gut microbiota, or the inflammatory process  
78 in general [4,17–20]. Moreover, familial aggregation of IBD has been suggested to be a risk  
79 factor for a more aggressive course of the disease, earlier disease onset, and greater need for  
80 treatment [21,22], while disease risk in families with IBD cases is generally increased [23]  
81 and can be further evaluated by polygenic risk scores derived from the accumulated genetic  
82 knowledge of the disease [24].

83 Most prior studies on IBD risk factors used a classical case–control design, comparing IBD  
84 patients to healthy controls. However, the molecular alterations or lifestyle factors that  
85 precede or predict the pre-clinical onset of IBD can only be identified reliably in prospective  
86 studies that collect health-related information and biomaterials from unaffected or preclinical  
87 individuals over time. Given the relatively low IBD incidence in the general population and  
88 the high familial risk for IBD, a family-based prospective study represents a more efficient  
89 study design than a population-based study. However, due to the high costs of maintenance,  
90 infrastructure, participant commitment, and long run-time, this type of study design is rare but  
91 offers great potential [23]. Thus, in 2013, the Inflammatory Bowel Disease Family Cohort  
92 (KINDRED cohort), the so-called KINDRED cohort, was initiated in Kiel (Germany) to  
93 systematically collect disease-relevant data from IBD patients and their healthy relatives. The  
94 main objectives of the study were (i) to identify lifestyle factors and biomarkers that  
95 predispose patients to, or predict, the development of IBD and (ii) to characterize the long-  
96 term clinical course of IBD patients. To this end, comprehensive and high-quality biomaterial  
97 (blood, stool, hair) as well as detailed health-related, lifestyle, sociodemographic, and kinship  
98 information was collected from the participants of this study. Regular follow-ups are  
99 performed (approximately every 2 years), which are embedded into a long-term follow-up  
100 program for healthy relatives with respect to IBD onset.

101 One main driver of IBD has been identified in the microbial community inhabiting the  
102 intestinal tract. Various human population studies and animal experiments have identified a  
103 multitude of ways in which the microbiome varies and interacts with its host, eventually  
104 leading to and becoming indispensable for the development of intestinal inflammation [25].

105 One current hypothesis is that dysbiosis, a shift in the balance of a beneficial microbial  
106 community as a whole in favor of a proinflammatory community, eventually leads to and  
107 sustains intestinal inflammation via metabolic and immunologic alterations but not single taxa

108 [25]. These community shifts can result from disturbances introduced through environmental  
109 and lifestyle factors, diet, hygiene, medication, or even the host's genetic makeup [25–27].  
110 However, the development of inflammation is a nonlinear and often long-lasting process that  
111 can persist preclinical (perceived healthy), although inflammation may already fester [28].  
112 Analyzing early changes, particularly in microbial communities, and reducing the influence of  
113 treatment-related disturbances offers immense opportunities to trace the true pathological  
114 dynamics of the microbiota during the development of IBD. Thus, recent efforts in predicting  
115 IBD in healthy or preclinical individuals have successfully been employed to identify  
116 dedicated transcriptional [29], physiological [30], and microbial markers that are broadly  
117 associated with disease onset in IBD [31,32].

118

## 119 **RESULTS**

### 120 *Current status of the KINDRED cohort*

121 As of April 2021, the Kiel KINDRED cohort included 1497 IBD patients and 1813 initially  
122 non-affected family members, belonging to a total of 1372 families. The data of all study  
123 participants whose baseline questionnaires were double entered and quality-checked were  
124 extracted from the database in a data freeze in March 2021, comprising 2393 individuals  
125 (1321 IBD patients and 1072 non-affected relatives). The discrepancy in the total number of  
126 study participants is due to the time lag between study inclusion and the return of biomaterial  
127 and questionnaire data (plus quality check). The baseline characteristics of the healthy  
128 participants, stratified by age, and of the IBD patients, stratified by disease subtype (CD, UC,  
129 uIBD), are summarized in Table 1 and Suppl. Tables S3 & S4, respectively.

130 Available biosamples from the baseline collection were used to generate data for biomarker  
131 discovery, including genome-wide genotyping (Global Screening Array (GSA), version 2.0

132 (Illumina), N=1725). These data are complemented by fecal 16S rRNA-based microbiome  
133 profiles (N=1812), blood biomarker profiles (C-reactive protein, N=316; hemoglobin, N=49;  
134 mainly IBD patients), serum antibodies (ASCA IgA, N=781; ASCA IgG, N=780; anti-GP2  
135 IgA, N=781; anti-GP2 IgG, N=781), and fecal indicators of inflammation (calprotectin,  
136 N=1763; occult hemoglobin/haptoglobin status, N=1760; Bristol stool score, N=882). For a  
137 subset from the first and second follow-ups (N<sub>F1</sub>=648, N<sub>F2</sub>=539), 16S rRNA sequencing data,  
138 fecal calprotectin (N<sub>F1</sub>=519, N<sub>F2</sub>=147), hemoglobin/haptoglobin measurements (N<sub>F1</sub>=516,  
139 N<sub>F2</sub>=144), Bristol stool score (N<sub>F1</sub>=316, N<sub>F2</sub>=255) and blood biomarkers for inflammation  
140 (Hb: N<sub>F1</sub>=203, N<sub>F2</sub>=162; CRP: N<sub>F1</sub>=179, N<sub>F2</sub>=127) were also available. Our study set  
141 included 7 cases of IBD onset (four in the BL → F1 group, one in the F1 → F2 group, and  
142 two in the F2 → F3 group; see Figure 1A).

#### 143 ***Distribution of physiological and clinical biomarkers of inflammation in the KINDRED*** 144 ***cohort***

145 We detected several highly interesting patterns among the biomarkers of inflammation, but  
146 our analyses focused on samples from the baseline examination because of the large number  
147 of samples and biomarkers. In the baseline samples, the concentrations of the antifungal  
148 antibodies anti-*Saccharomyces cerevisiae* antibody (ASCA) IgG and IgA decreased with age  
149 and increased with BMI (Table 2). ASCA antibody levels were greater in CD patients than in  
150 healthy controls and UC patients (Figure 1B). Furthermore, the concentration of anti-  
151 glycoprotein 2 IgA and IgG (GP2), which have previously been reported to be reliable  
152 biomarkers for PSC and CD [33–35], significantly differ between healthy relatives and IBD  
153 patients, and are significantly associated with patient age, BMI and sex, with the  
154 concentration of anti-GP2 Ig's decreasing in females and increasing in males. We detected a  
155 particularly high average concentration in CD patients compared to either that in UC patients  
156 or non-IBD control subjects (Table 2). The fecal calprotectin concentration, a general measure  
157 of inflammatory processes in the gut and a metal sequestering protein, decreases with age in

158 IBD patients but significantly increases in CD patients, followed by UC patients (Figure 1B).  
159 The concentration of C-reactive protein (CRP), similar to hemoglobin (Hb) levels, both  
160 measured only in diseased individuals, increased with BMI, while Hb levels in affected males  
161 were significantly greater than for women with IBD (Figure S4). Stool consistency, as  
162 expressed by the self-reported Bristol stool score (BSS), strongly and significantly increases  
163 with age. This age-related increase in BSS was particularly strong in CD patients (Figure 1B;  
164 Table 2). This indicates that older CD patients suffer from greater bowel irritation and thus  
165 shorter intestinal passage times. At other examination time points, the associations were  
166 mostly comparable to the associations in the baseline analyses, as reported in Table 2.

167 ***Genetic predisposition to IBD can be expressed by polygenic risk scores (PRS) and is***  
168 ***associated to physiological signs of inflammation***

169 Due to the increasing number of variants associated to IBD it has become possible to quantify  
170 an individual's genetic risk or genetic predisposition to CD, UC, or IBD using Polygenic Risk  
171 Scores (PRS) [36]. On average CD- and UC-PRS were relatively disease specific as they only  
172 showed significant differences from healthy KINDRED participants in the respective patient  
173 subset, with the exception of the more general IBD-PRS which was clearly increased in CD  
174 and UC patients alike, compared to healthy control individuals (Figure 1C). Interestingly, the  
175 average UC-PRS and general IBD-PRS were significantly elevated among future onset cases,  
176 while the CD-PRS showed no clear pattern. In addition to the average higher genetic risk  
177 present in future onset cases, the probability of IBD onset was significantly associated with  
178 UC-PRS and IBD-PRS, even though only nominally statistically significant (see Table S6,  
179 Figure S6A). The genetic risk for IBD should naturally be increased in IBD patients,  
180 compared to healthy individuals in general, a pattern displayed by all three PRS variants.  
181 However, closely related healthy first-degree relatives (FDRs) showed no significantly  
182 elevated PRS. The average genetic risk for IBD did not significantly differ among healthy

183 FDRs, more distant relatives (>1st degree relatives), and unrelated controls (Figure S7). This  
184 pattern points to a significantly stronger accumulation of risk variants in affected individuals  
185 and onset patients, even in comparison to closely related family members and a high disease  
186 specificity of the respective PRS except for the more general IBD-PRS.

187 In our analyses, we tested PRS-CD, PRS-UC, and PRS-IBD for associations with different  
188 biomarkers. We detected a significant positive relationship only between CD- and IBD-PRS  
189 and the levels of ASCA-IgA/IgG and, to some extent, between IBD-PRS and the levels of  
190 anti-GP2-IgA (Figure 1D, Figure S5, Table S5). However, other inflammatory markers, such  
191 as calprotectin, BSS, CRP, and Hb did not show any, or only weak associations with genetic  
192 risk scores for IBD (Figure 1D, Figure S5, Table S5).

### 193 ***Differential abundance of microbial taxa across IBD pathologies and time***

194 Among the three time points we investigated, the fecal communities were dominated mainly  
195 by members of the phylum Bacteroidetes, followed by members of Firmicutes and  
196 Proteobacteria, and minor taxonomic groups (*e.g.* Fusobacteria, Tenericutes, Candidatus  
197 Saccharibacteria; Figure 2A, Figure S8). At this high taxonomic level strong differences  
198 between the main IBD pathologies are detectable, as reflected in the association of  
199 Bacteroidetes and Firmicutes with the control individuals and the associations of  
200 Proteobacteria, Fusobacteria and Cand. Saccharibacteria with IBD, particularly CD (Figure  
201 S6, Table S7). These associations are consistent over time and carry down to the ASV level  
202 (see below).

203 By investigating abundance differences at the ASV level, we detected more than 750  
204 significant differentially abundant taxa/ASVs among the main IBD pathologies (BL: 294, F1:  
205 150, F2: 193, combined: 536; Figure 2B). At the baseline time point we detected taxa  
206 belonging to the *Enterobacteriaceae* (Gammaproteobacteria), such as *Haemophilus*,



207 *Escherichia/Shigella*, *Raoultella*, and *Klebsiella*, which were strongly overabundant in IBD  
208 patients (Figure 2C). Particularly in the fecal community of CD patients we observed a high  
209 abundance of *Klebsiella* and other *Enterobacteriaceae*, which persisted over time (see Figure  
210 2C, S9, S10). In the phylum Firmicutes we found various associations with different IBD  
211 pathologies. Members of *Veillonella*, *Megasphaera*, *Dialister*, *Enterococcus*, and  
212 *Granulicatella* were more abundant in CD and UC patients alike. In contrast, members of the  
213 *Ruminococcaceae* (e.g. *Clostridium IV*, *Oscillibacter*, *Faecalibacterium*) are consistently  
214 overrepresented in healthy individuals. *Erysipelotrichaceae* is more abundant in IBD patients  
215 and is frequently associated with UC. Members of the *Bacteroidetes*, with the exception of  
216 some *Bacteroides sp.* are consistently more abundant in healthy individuals, e.g. *Alistipes*,  
217 *Prevotella*, *Alloprevotella*, *Parabacteroides*, and *Bacteroides* (see Figure 2C, S9, S10).

218 Interestingly IBD-associated taxa belonging to *Veillonella*, *Dialister* and *Granulicatella*  
219 (Firmicutes), but also *Haemophilus*, *Klebsiella* (Proteobacteria), and *Rothia* (Actinobacteria)  
220 are prominent indicators of oral colonization of the intestinal microbiome. Similarly, a rarely  
221 described disease associated taxon, *Cand. Saccharibacteria* [37] are commonly found in the  
222 oral microbiome. *Fusobacteria* (i.e. *F. nucleatum*), which are generally associated with the  
223 development of colorectal cancers, were found to be overabundant in CD patients and could  
224 also be replicated in external cohorts (see Table S8-S11). The majority of associations, such  
225 as the associations of *Veillonella*, or members of *Clostridium XIVa clostridioforme*),  
226 *Clostridium XVIII ramosum*), and *Clostridium sensu stricto* groups with IBD, or the higher  
227 abundances of ASVs belonging to *Alistipes* and *uncl. Clostridiales* in healthy individuals, can  
228 be consistently detected in external cohorts, although some incongruities exist between  
229 cohorts (e.g. *Oscillibacter* ASVs in KINDRED-Contr./SIC-IBD patients).

230 Various taxa showed significant correlations with clinical indicators of intestinal  
231 inflammation and genetic risk scores. Indicators of oralization correlate with physiological

232 signs of inflammation (e.g. *R. mucilaginosa*-BSS ↓, *Cand. Saccharibacter*-calprotectin ↑,  
233 *Klebsiella uncl.*- BSS ↑, *F. nucleatum*-calprotectin ↑, *Dialister invisus*-BSS↓ & anti-GP2  
234 IgA/IgG ↓, *Veillonella uncl.*-IBD-PRS ↑). Interestingly, most Firmicutes correlated very  
235 consistently and almost entirely positively with the Bristol stool score (BSS) under IBD  
236 conditions (IBD, CD, UC), while members of the Bacteroidetes correlated very consistently  
237 negatively with BSS, encompassing pathobionts and commonly probiotic taxa (see Figure 2C,  
238 Table S12). This pattern was also present in healthy individuals, but did not reach  
239 significance. When we focused on the association of bacteria with genetic risk for IBD, we  
240 found that most Bacteroidetes (e.g. *Alistipes*) were correlated with decreasing genetic risk for  
241 CD, UC, and IBD, with the exception of *Bacteroides fragillis* (Figure S11, Table S13).  
242 Interestingly, *Barnesiella uncl.*, a common butyrate producer, correlated positively with  
243 genetic CD risk in healthy individuals, but correlated negatively with genetic CD risk in  
244 diseased individuals. In general, correlations with clinical parameters and genetic risk were  
245 most frequent in IBD patients, especially in CD patients, while almost no significant  
246 correlations were detected in UC patients. Different IBD associated Proteobacteria,  
247 specifically members of *Enterobacteriaceae* are strongly associated with increased ASCA  
248 IgG/IgA levels and calprotectin levels, as well as decreased gut transit time (increased BSS;  
249 e.g. ASV\_99 *Escherichia/Shigella coli*) only in IBD/CD patients, similar to the findings for  
250 the IBD/CD associated Firmicutes *Clostridium XIVa clostridioforme* (ASV-103). Moreover,  
251 the *Enterobacteriaceae* (e.g. ASV-42-*Enterobacteriaceae uncl.*, ASV-165-*Citrobacter uncl.*),  
252 as well as members of the *Clostridium XIVa* strongly correlated with increasing genetic risk  
253 for IBD, and more precisely CD in the patient cohorts (Figure 2C & Figure S11; Table S12).  
254 Several Firmicutes SCFA producers correlate with decreasing ASCA IgG/IgA levels and IBD  
255 risk in healthy and diseased individuals (e.g. ASV\_71 *Faecalibacterium uncl.*), while the  
256 abundances of others correlate with elevated levels of genetic risk, particularly in diseased  
257 individuals (e.g. *Blautia*, *Ruminococcaceae uncl.*). Overall, we observed consistent

258 associations of ASVs to genetic risk for IBD and different clinical markers for IBD, with only  
259 minor associations in UC patients or with genetic risk for UC.

260 *Alpha diversity patterns associated with IBD pathology, clinical characteristics, and*  
261 *microbial dysbiosis*

262 A critical characteristic of ecosystems is their complexity, which can be informative for  
263 assessing the community state, productivity or stability and can be based on, e.g. the number  
264 of community members, their distribution, or their phylogenetic relationships. Overall, we  
265 identified an extensive loss of species diversity in IBD patients with respect to species  
266 richness (Figure 2D), complexity or phylogenetic relatedness of community members  
267 (NTI<0, NRI<0) (Figure S12, Table 3). Individuals suffering from CD show a significantly  
268 greater loss of diversity than UC patients based on richness and complexity but not at the  
269 phylogenetic level. Onset cases only showed a slight but insignificant reduction in diversity  
270 (less richness, less complexity, less phylogenetic dispersion).

271 Based on the differential abundance patterns we derived two indices of host health based on  
272 the bacterial abundance and distribution, namely, the Microbial Dysbiosis index (MD-index)  
273 as introduced by Gevers *et al.* (2014) and the more generalized GMHI (general microbiome  
274 health index) as developed by Gupta *et al.* (2020) [38,39]. Focusing on the MD-index we can  
275 observe a clearly negative correlation of alpha diversity with community dysbiosis at any  
276 level, further implying an accelerated loss of taxon diversity with increasing dysbiosis  
277 (breakpoint at MD=0.615, Figure 2D, Table S13). The phylogenetic composition changes to  
278 become more random with increasing dysbiosis but shows a less severe pattern than species  
279 richness or abundance distribution (Figure S12, Table S13). Across all time points we can  
280 identified a trend of strongly decreasing diversity in UC and CD patients with increasing  
281 dysbiosis, while healthy controls showed a far slower decline of diversity under increasing  
282 dysbiosis, particularly in the taxon-based indices (Table S14). The correlations of different

283 alpha diversity metrics with the GMHI showed comparable but weaker relationships as  
284 observed with the MD-index. Alpha diversity metrics are further negatively (Chao1, Shannon  
285 H) and positively (NRI, NTI) correlated with physiological signs of inflammation, such as  
286 ASCA-IgA/IgG, calprotectin, and BSS, particularly in IBD patients, which translates to more  
287 severe signs of inflammation at a lower community diversity in diseased individuals (CD, UC,  
288 uIBD; Figure 2G, Table S15).

289 *Community differences between IBD pathologies are associated with clinical-,*  
290 *anthropometric- characteristics, patterns of community dysbiosis, and genetic risk*

291 By investigating community similarities at different levels, we revealed various patterns of  
292 health-associated, anthropometric, and lifestyle related community differences. In particular,  
293 IBD pathology showed a strong association with community differences, with the strongest  
294 separation between microbial communities originating from healthy individuals and those  
295 originating from CD patients (Figure 3A; Table S16). In this context, we observed a strong  
296 association of the taxonomic community composition with the MD-index and to a lesser  
297 extent with GMHI (Table S16). This gradient of dysbiosis parallels the community  
298 differentiation between healthy and diseased individuals, but also recapitulates significant  
299 gradients of physiological markers for inflammation such as BSS, ASCA IgA/IgG levels, and  
300 calprotectin levels in the communities (see Figure 3A, 3B, 3J; Figure S13, S14; Table S17 &  
301 Table S18). We also detected significant, yet weak correlations between taxonomic  
302 community composition and polygenic risk scores, which were very weak for UC associated  
303 genetic risk and strongest for CD and general IBD risk at the taxonomic- (Figure S7D) and  
304 functional levels (Figure S7E) and followed the direction of the dysbiosis gradient. Similarly,  
305 the cumulative genetic risk for CD and general IBD positively correlated with the MD-index,  
306 while UC-PRS displayed no significant correlation, similar to GMHI (Figure S5B & S5C).

307 Furthermore, onset cases were mostly distributed around the mean of the community  
308 compositional distribution and dysbiosis gradient, and thus did not represent community  
309 outliers (Figure 3C, Figure S13 & S14). The MD-index (and GMHI) strongly distinguished  
310 IBD and non-IBD individuals and corresponded with the clustering of microbial communities  
311 with respect to IBD pathology, even if applied to external patient cohorts (Sweden, Malta)  
312 (Figure 3E-3H, Figure S14G). The MD-index is most strongly elevated in CD patients and  
313 further distinguished CD patients (particularly treatment-naïve) from healthy controls in the  
314 Maltese cohort (Figure 3H). GMHI displayed on average a more consistent decrease in  
315 diseased individuals across all cohorts, and further distinguished UC patients from healthy  
316 controls (Figure S14H). These “universal” disease patterns are also reflected in the sample  
317 clustering by IBD condition, irrespective of cohort origin (conditional dbRDA (conditioned  
318 by geographic origin):  $F_{4,3308}=4.0406$ ,  $P<0.0001$ ,  $adj. R^2=0.0037$ ) and through the significantly  
319 elevated taxonomic and functional community variability in IBD patients compared to healthy  
320 controls within (Figure 3D, Figure S14H & S15, Table S20) and across cohorts (Figure 3G &  
321 S14H).

322 Both dysbiosis scores corresponded closely with several physiological signs of inflammation,  
323 such as the reactivity to fungal antigens as expressed via ASCA IgG or IgA titers, or related  
324 immunological markers, like anti-GP2 IgG and IgA levels (Figure 3I & 3J, Figure S16, Table  
325 S18). ASCA antibody levels increased most strongly in CD patients with increasing MD score  
326 (decreasing GMHI). In contrast, anti-GP2 IgA/IgG levels positively correlate with MD but  
327 appear to show a slight overall increase and higher average MD levels in CD and UC patients,  
328 as expected. Overall, the relationship of GMHI and MD-index with ASCA or anti-GP2  
329 antibody levels were comparable (Table S18). The GMHI and MD-index displayed strong  
330 associations with the BSS, with lower stool consistency in IBD patients and increasing  
331 dysbiosis (MD ↑, GMHI ↓) (Figure 3I). The MD-index was strongly associated with

332 increasing BMI and calprotectin levels, which were on average higher in CD and UC patients  
333 than in healthy controls (Figure 3I, Figure 1B). The concentrations of serum CRP- and Hb  
334 levels, which were only assessed in diseased individuals, were not significantly associated  
335 with microbial dysbiosis metrics (Figure S16). The practitioner's overall assessment of  
336 disease severity in IBD patients only (1-no/4-severe inflammation), revealed a strong positive  
337 association with MD, while GMHI was not significantly correlated.

338 When we investigated dysbiosis patterns among FDRs, we observed that the MD and GMHI  
339 in affected individuals were greater or lower, respectively, than those in healthy family  
340 members. However, healthy FDRs of IBD patients, generally related, or unrelated healthy  
341 individuals, did not show a greater degree of dysbiosis. However, the MD-index was weakly  
342 associated with an increased probability of IBD onset ( $DF=1,1735$ ,  $Z=-2.0685$ ,  $P(Z)=0.0386$ ,  
343  $N_{onset}=7/1815$ ;  $DF=1,726$ ,  $Z=-1.2579$ ,  $P(Z)=0.2084$ ,  $N_{onset}=7/794$ ; binomial GLM), while the  
344 GMHI showed no significant association with IBD onset.

345 IBD pathologies were consistently the most influential health conditions for the taxonomic  
346 and functional differences of the microbial communities (Figure 3J & Figure 4D, Table S21).  
347 The most pronounced differences were between the healthy controls and CD patients,  
348 followed by the differences between healthy controls and UC patients, while functional and  
349 taxonomic differences between UC- and CD patients persist (see Table S16, Table S21).  
350 Together with the presence of a stoma/artificial bowel outlet (colostomy, ileostomy,  
351 permanent/temporary stoma), other pathologies that influence the anatomic structure or gut  
352 transition time, such as obstipation, diarrhea, and intestinal stenosis strongly influence the  
353 taxonomic and functional composition of the microbial communities (Figure 3J). Thus,  
354 chronic and acute diseases, which influence passage time, have a significant impact on the  
355 taxonomic and functional composition of the intestinal microbiome. This pattern remained

356 consistent even after excluding potential confounding effects such as age, BMI, sex, and IBD  
357 pathology (see Table S21).

358 In addition to the impact of different pathologies, most importantly IBD, we observed a  
359 significant impact of various pharmaceutical treatments on microbial communities.  
360 Interestingly, antibiotic treatments between 6 weeks and one year prior had significant but not  
361 prominent effects as reported by [40]. However, the most influential were drugs that modify  
362 the fecal passage time (*e.g.* Loperamide, surfactants), further pronouncing the strong and  
363 consistent association of BSS with community composition. In particular, these  
364 pharmaceuticals remained influential after correcting for IBD condition, age, BMI, and sex  
365 (Figure 3J, Table S22) and associated with higher levels of BSS ( $W=24756$ ,  $P=3.3910\times 10^{-9}$ ,  
366 Wilcoxon-test).

367 Investigating the impact of macro and micronutrient intake on the microbiota revealed a very  
368 interesting parallel distribution of F20:4 (Eicosatetraenoic acid/Arachidonic acid) and  
369 MD/GMHI indices with the microbial community composition (Figure 3J, Figure S13G-I),  
370 which was the only nutrient whose intake associated with an increasing dysbiosis signal in the  
371 microbial community, in direct correlation (Figures S14J; Table S23), and consistent in  
372 follow-up time points (see Figure 3J & S13G-I; Table S23). This eicosanoid precursor can be  
373 a potent, immunologically active molecule [41] and its uptake at baseline was weakly  
374 correlated to MD ( $\rho=0.0486$ ,  $P=0.0772$ , Spearman correlation), BMI ( $\rho=0.1278$ ,  
375  $P=4.7168\times 10^{-6}$ ), ASCA IgG ( $\rho=0.0647$ ,  $P=0.0881$ ), and calprotectin ( $\rho=0.0514$ ,  $P=0.0648$ ).  
376 Increasingly dysbiotic communities were further correlated to increasing consumption of  
377 glycogen, various amino acids and general protein uptake (ZE), which can be interpreted as  
378 elevated uptake of animal products (Figures S14J), which paralleled the gradients of  
379 inflammatory markers (*e.g.* calprotectin; see Figures S13D-F & Figures S13G-I). In contrast,  
380 the increased uptake of various vitamins and trace minerals, as well as long-chain

381 carbohydrates (e.g. ZB-Fiber) significantly coincided with a less inflammatory/disturbed  
382 community structure (GMHI and MD) and were associated with reduced physiological signs  
383 of inflammation, potentially via SCFA production [42]. The functional community  
384 composition was only associated with vitamin B derivatives and trace mineral consumption  
385 (see Figure 4D & Figure S17G-I; Table S23).

386 ***Uncharacteristic community clusters (DMM) are associated with dysbiosis and clinical***  
387 ***signs of inflammation***

388 Unsupervised community clustering revealed three community clusters, which do not fully  
389 correspond to previous reports [43], but share common indicators reported in the literature  
390 (Figure 4A). Cluster-1 is characterized by a high abundance of *Prevotella* and  
391 *Ruminococcaceae* (*Prevotella*:  $P_{FDR}=1.1162\times 10^{-26}$ ; *Ruminococcus*:  $P_{FDR}=2.1846\times 10^{-59}$ ;  
392 Kruskal test), while cluster-2 shows a high abundance of e.g. *Bacteroides* and  
393 *Faecalibacterium* (*Bacteroides*:  $P_{FDR}=1.4604\times 10^{-17}$ ; *Faecalibacterium*:  $P_{FDR}=1.5310\times 10^{-84}$ ;  
394 Kruskal test). In contrast, individuals in Cluster-3 showed a greater propensity for  
395 inflammation as indicated by the elevated MD-index ( $P<2.2\times 10^{-16}$ ; Kruskal test, Figure 4B &  
396 4C) and community variability (Figure 4D), as well as increased levels of physiological  
397 inflammation markers such as ASCA-IgG/IgA, and increased stool moistness/softness (Figure  
398 S18A). This cluster further showed a higher abundance of *Enterobacteriaceae* (e.g.  
399 *Escherichia/Shigella*:  $P_{FDR}=4.0936\times 10^{-61}$ ; Kruskal test), which is unconventional for so-called  
400 “enterotypes”. However, this community type has been reported before and appears to  
401 correspond to a dysbiotic community state, positioned outside the standard community states  
402 reported so far [44–46]. In addition, we detected a significantly greater level of genetic risk for  
403 CD ( $F_{2,1665}=12.5869$ ,  $P=3.7541\times 10^{-6}$ ; LM) and general IBD ( $F_{2,1665}=4.3156$ ,  $P=0.0135$ ; LM)  
404 among individuals in this unconventional community cluster, but not for UC risk



405 ( $F_{2,1665}=0.1712$ ,  $P=0.8427$ ; Figure S18B), indicating the strong effect of IBD and a broad  
406 genetic disease risk on community composition.

407 ***Imputed community functions reveal the importance of amino acid related pathways and***  
408 ***bacterial surface structures in health and disease***

409 On average, communities differ more strongly between IBD pathologies based on the imputed  
410 functions compared to taxonomic differences, particularly regarding the difference between  
411 healthy controls and CD patients (Table S16). The ranking of single associations of  
412 pathologies, physiological parameters, and pharmaceutical treatments with functional  
413 community differences overlaps with the taxon-based analyses, although fewer associations  
414 can be detected (Figure 4E & 4F, Figure S17, Table S17, S21, S22, S23). Additionally,  
415 physiological signs of inflammation, such as ASCA IgA/IgG or decreasing stool consistency  
416 parallel the dysbiosis gradient in the functional composition (Figure 4E, Figure S17, Figure  
417 S17D; Table S16, S17). The MD-index and GMHI explained a greater percentage of  
418 community variation at the functional level (see Figure 4A, Figure S17M; Table S16), while  
419 functional community differences were only weakly correlated with the genetic risk scores for  
420 CD ( $P<0.05$ ; Figure S6E).

421 At the level of single functions, we detected various differentially abundant KEGG  
422 Orthologous Groups (BL:  $N_{total}=5071$ , F1:  $N_{total}=4347$ ; F2:  $N_{total}=4402$ ). By investigating  
423 systematic changes in functional abundances via Gene Set Enrichment Analysis [47], we  
424 identified various highly persistent functional differences between disease states, which were  
425 further supported in the external cohorts (Table S24). Various metabolic pathways involved in  
426 the metabolism and biosynthesis of amino acids, ranging from general *Biosynthesis of amino*  
427 *acids*, to *Phenylalanine, tyrosine and tryptophan biosynthesis*, among others, are consistently  
428 enriched in healthy individuals or rarely in UC patients compared to CD patients. Other amino  
429 acid-related pathways were similarly enriched in healthy individuals (*Arginine biosynthesis*;

430 *Alanine, aspartate and glutamate metabolism; Valine, leucine and isoleucine biosynthesis;*  
431 *Lysine biosynthesis*; see Figure 4E; Table S24). Basic metabolic functions and cell growth  
432 were more strongly represented in healthy controls and were repeated among the other cohorts  
433 in a similar direction (e.g. *Citrate cycle (TCA cycle), Oxidative phosphorylation*; Figure 4H,  
434 Figure S19-S22; Table S24), as were pathways involved in bacterial motility and biofilm  
435 formation, which were enriched in healthy individuals (*Bacterial chemotaxis, Flagellar*  
436 *assembly, Biofilm formation - Pseudomonas aeruginosa*). Oxidative phosphorylation also  
437 entails anaerobic electron transfer reactions (anaerobic respiration) that reduce other electron  
438 acceptors than oxygen for ATP generation (e.g. nitrate, sulfate, fumarate), as performed by  
439 various intestinal bacteria (e.g. *Desulfovibrio*). In contrast, functions related to carbohydrate  
440 degradation (e.g. *ABC transporters, Phosphotransferase system (PTS), Fructose and mannose*  
441 *metabolism, Galactose metabolism*) were very consistently enriched in IBD-associated  
442 communities and may, in combination with changing SCFA levels (F1 & F2: *Propanoate*  
443 *metabolism and Butanoate metabolism*), contribute to mucosal degradation. Bacterial traits  
444 that are central to the generation and integration of immune modulatory factors of gram-  
445 negative (*Lipopolysaccharide biosynthesis*) and gram-positive bacteria (*Teichoic acid*  
446 *biosynthesis, Peptidoglycan biosynthesis*) were also consistently enriched among IBD patients  
447 (Figure S19-S22; Table S24).

448 ***Network analyses show systemic and transferrable patterns of community disturbance, and***  
449 ***high disease dependent taxon centrality in community networks***

450 To analyze the microbial community as a whole, we constructed co-abundance networks to  
451 identify important taxa and patterns in the fecal microbial communities that may be central for  
452 community and host homeostasis (Figure 5A). Across networks (split by time point, or by  
453 time point and pathology), we observed significant differences in the composition and  
454 topography at the network level between IBD pathologies and healthy community networks in

455 the KINDRED cohort (control vs. IBD:  $F_{1,7}=2.4873$ ,  $P=0.0947$ ,  $R^2=0.2622$ ,  $adj. R^2=0.1568$ ;  
456 see Figure 5B). This also held true when external IBD cohorts were included (Control vs.  
457 IBD:  $F_{1,12}=2.1301$ ,  $P=0.0604$ ,  $R^2=0.1507$ ,  $adj. R^2=0.0800$ , Control vs. IBD (incl. general IBD  
458 networks):  $F_{1,18}=2.4144$ ,  $P=0.0412$ ,  $R^2=0.1183$ ,  $adj. R^2=0.0693$ , Graphlet distance; Figure 5C  
459 & Figure S27A). Furthermore, we observed signs of network/community contraction in IBD  
460 communities in the KINDRED cohort, as implied by the increasing network density and  
461 radius in diseased community networks, which remained consistent when we included  
462 networks derived from healthy and diseased external cohorts (Figure S27B). In addition,  
463 “natural connectivity” is increased in IBD communities, which strongly suggests a greater  
464 resistance to disturbances or stochastic changes by those communities. A higher “natural  
465 connectivity” after disease manifestation implies more densely connected networks with  
466 fewer bottlenecks, and pronounced community structure, which tend to possess a greater  
467 robustness against random fluctuations and community decay [48,49].

468 At the single taxon level/node level (Figure 5A, Figure S23-S26), we observed various  
469 pathogenic and IBD associated groups, consisting mainly of *Clostridium XVIII ramosum*  
470 (*Erysipelotrichaceae*: ASV\_193) and *Clostridium XIVa clostridioforme/spec.* (e.g. ASV\_103,  
471 ASV\_171), as well as various members of the oral community (e.g. *Veillonella*, *Rothia*, *Cand.*  
472 *Saccharibacteria*, *Fusobacterium*) and *Enterobacteriaceae* (e.g. *Klebsiella*), to be consistently  
473 central across time points (particularly in diseased networks) and characteristically abundant  
474 in IBD patients. Similarly *Flavonifractor plautii* (ASV\_93, *Ruminococcaceae*) or the  
475 Actinobacterium *Eggerthella lenta* (ASV\_713), are elevated as well in IBD, highly central,  
476 and known for pro-inflammatory characteristics [50–54] (Figure S23-S26, Table S25). In  
477 contrast, ASVs belonging to *Alistipes* (i.e. *A. putredinis*, *Alistipes spec.*, *A. shahii*) are highly  
478 central and consistently abundant in control individuals across time points and external  
479 cohorts, like other SCFA producing bacteria (e.g. *Porphyromonadaceae*: *Barnesiella*,

480 *Odoribacter*, *Parabacteroides spec.*; *Lachnospiracea: Clostridium XIVb uncl.*;  
481 *Ruminococcacea (Clostridial cluster IV, Ruminococcus*; Figure 5A, Figure S23-S26, Table  
482 S25). However, through the nature of network construction through statistical means,  
483 centralities, and key-stone status may vary [55,56].

#### 484 ***Heritability of microbial taxa***

485 Using a mixed model framework, we investigated the heritability of single ASVs using  
486 pedigree relationships in the cohort. Most of these associations were driven not only by host  
487 kinship or family conditions, but also by environmental/ anthropometric/ lifestyle  
488 characteristics (age, sex, BMI, and IBD status). At the ASV level, the taxa showing the  
489 highest heritability estimates were mainly members of the *Bacteroides*, while particularly  
490 *Bacteroides uniformis* (ASV-6,  $h^2_{null}=0.4723$ ) and *Paraprevotella uncl.* (ASV 84,  
491  $h^2_{full}=0.4783$ ) had the highest heritability estimates (Table S26). Most taxa were impacted by  
492 environmental influences in addition to the apparent kinship effect/familiar transmission  
493 ( $h^2_{null} < h^2_{full}$ ; Figure 5D, Figure S28A; Table S26). Interestingly, potential probiotic bacteria  
494 abundant in healthy individuals, or mutualistic taxa (not associated to IBD) displayed high  
495 heritability (e.g. *Parabacteroides uncl.*, *Feacalibacteria uncl.*, *A. muciniphila*, *B.*  
496 *thetaiotaomicron*, *Alistipes uncl.*). *Akkermansia muciniphila* displayed a lower impact of  
497 environmental/anthropogenic variables ( $h^2_{null} > h^2_{full}$ ) on the inheritance patterns, and displayed  
498 a better fit solely by relatedness ( $h^2_{null}=0.2573$ , Figure 5D, Figure S28A), similar to other  
499 highly heritable taxa, such as the ubiquitous human gut commensal *Bacteroides uniformis*  
500 [57] (Table S26). However, among the top 25% percent of the most heritable ASVs, we  
501 observed disease-associated taxa, mainly belonging to the Proteobacteria such as e.g.  
502 *Escherichia/Shigella* (ASV-5, ASV-10) or the typical oral *Dialister invisus* (ASV-17,  
503 Firmicutes), which were more abundant in IBD patients. The community wide dysbiosis  
504 indices (MD-index and GMHI), showed relatively strong heritability as well, which implies a

505 genetic influence or intra-family transmission of dysbiosis. GMHI was within the boundaries  
506 of ASV based estimates in our cohort ( $h^2_{full}=0.1401$ ), while the heritability of the MD-index  
507 was clearly lower ( $h^2_{full}=0.0813$ ) and distinctly lower than estimates reported by Turpin *et al.*  
508 2016 ( $h^2=0.2728$ ) for the MD-index or the heritability estimates for IBD itself (CD: 0.25-0.42,  
509 UC: 0.04-0.15) [58,59]. Overall, heritability of the communities appears to be disturbed in  
510 cases of IBD, as we observed a reduced correlation of community composition and genetic  
511 distance (kinship, IBS distance) in IBD cases compared to healthy individuals (Figure S28B),  
512 although overall correlations were still significant (kinship-  $r=0.4406$ ,  $m^2_{12}=0.8059$ ,  
513  $P=0.0009$ , IBS-  $r=0.4272$ ,  $m^2_{12}=0.8175$ ,  $P=0.0009$ ; Procrustes test).

514

## 515 **DISCUSSION**

516 In this cohort study we investigated the state and evolution of IBD by combining taxonomic  
517 and functional microbial data from several successive time points with extensive  
518 anthropometric, medical, nutritional, and social information. Genome-wide association  
519 studies focusing on IBD revealed various genetic risk loci for IBD, which may contribute to  
520 immune dysregulation, antimicrobial immune reactions, and changes in host-microbiome  
521 homeostasis [60,17]. Various studies have investigated the interplay between microbes,  
522 metabolites, and host characteristics. These analyses revealed distinct physiological and  
523 microbial manifestations of the different IBD pathologies, which are reflected in external  
524 disease cohorts. The unsupervised community clustering (DMM community cluster 3) and  
525 significant differences of IBD patients across cohorts, highlight the severity of community  
526 turnover during IBD [46] and a strong signal of dysbiosis which may even warrant one or  
527 multiple distinct community states. Similarly, the strong differentiation of IBD-specific  
528 networks from healthy ones points to large changes in community dynamics after disease  
529 manifestation. These changes potentially result in a resilient or resistant community state as

530 implied by the more connected, denser, and robust networks after IBD manifestation (radius,  
531 density, natural connectivity) even across cohorts [49]. The consistently observed increased  
532 community variability in diseased individuals is a probable result of community transitions.  
533 This phenomenon, the so called “Anna Karenina pattern”, describes a pattern in which  
534 microbial communities of dysbiotic/diseased individuals are more different from each other  
535 than communities of healthy individuals are among each other. This phenomenon may be  
536 attributed to a loss of regulation or strongly individualized stochastic effects on community  
537 composition after IBD development [61] and has been detected in other IBD studies [25,62–  
538 64]. In summary, different community characteristics, such as increasing interindividual  
539 variability, greater community density/connectivity, and large compositional changes after  
540 IBD manifestation, point toward large and to some extent universal community  
541 transformations in IBD development.

542 ***Physiological and genetic markers of inflammation are strongly disease specific, associated***  
543 ***to the microbiome, and associate to disease onset***

544 The two main IBD pathologies, CD and UC, are physiologically, anatomically, and  
545 microbiologically quite distinct. We identified disease specific patterns in several  
546 physiological markers of IBD. In particular, anti-fungal ASCA IgG and IgA have strong  
547 associations to CD and IBD in general. ASCA Ig levels were not reported to be associated to  
548 high risk genetic markers of CD (CARD15/NOD2), but that its effect sizes might be  
549 genetically determined [65]. Thus, ASCA levels themselves are not genetically determined,  
550 but rather their responsiveness, which reflects the strength of inflammatory reactions to  
551 environmental antigens, potentially also of bacterial origin. In contrast we showed a  
552 significant association with polygenic risk scores for IBD, particularly CD, which was not yet  
553 shown before (Lee et al., 2021), and might originate from the wider array of cumulative

554 genetic risk factors known since the investigation by Halfvarson et al., 2005. This strong  
555 relationship of the PRS was not present for other physiological IBD biomarkers in this study.

556 Similar to ASCA Ig levels, anti-glycoprotein 2 antibodies (GP2) have been described as  
557 strong biomarkers for IBD [33,35], while fecal calprotectin represents a biomarker of general  
558 intestinal inflammation. All displayed strong disease specific differences and are strongly  
559 correlated with bacterial community characteristics such as composition, diversity, dysbiosis  
560 indices, but far weaker associated to the genetic risk scores for IBD. Bacterial taxa that  
561 associated to ASCA levels overlap only in parts with a recent study in UC patients at the  
562 family level (*e.g. Ruminococcaceae*) [66].

563 Decreased stool consistency, an indicator of reduced intestinal transit time, elevated fecal  
564 water content and eventually diarrhea, are hallmarks of IBD pathology. Although not  
565 associated with genetic risk for IBD, stool consistency decreases dramatically with increasing  
566 dysbiosis, diversity, and community composition. Recently Vandeputte et al. showed the  
567 effects of fecal passage time on community patterns in a Belgian cohort, surpassing dietary  
568 and disease related signals [67,68]. We found similar patterns, with increasing passage time  
569 (lower BSS) associated with increasing populations of Bacteroidetes and decreasing  
570 abundances of Firmicutes members. These patterns have mostly a similar orientation, but are  
571 significantly exaggerated in IBD patients, which implies a greater impact of bacteria on  
572 passage time/stool consistency or a greater impact of passage time on the bacterial community  
573 under intestinal inflammation. In line with the centrality of stool consistency and transit time  
574 for community composition, we observed a strong gradient of community composition and  
575 IBD associated dysbiosis with decreasing transit time. Similarly, among the most influential  
576 pharmaceutical treatments in the cohort are anti-diarrhea drugs, and the most influential  
577 medical conditions in addition to IBD are conditions affecting fecal flow (stoma, obstipation,  
578 diarrhea, stenoses, etc.). Changes in transit time can change the taxonomic and functional

579 composition-, bidirectionally interact with the microbiota, and alter the metabolite  
580 composition by transitioning from saccharolytic to proteolytic metabolism [69,70].

581 We determined a clear connection between the genetic risk of IBD development, particularly  
582 CD and general IBD, and the microbial community composition or level of dysbiosis. This  
583 finding emphasizes the potential connection between genetic risk factors such as *e.g.* *FUT2* or  
584 *NOD2* [60,71] and their combined influence on the microbial community, potentially leading  
585 to dysbiotic community dynamics in interaction with environmental factors. Polygenic risk  
586 scores integrate the abundance and penetrance of known risk variants for diseases such as  
587 IBD into a single genome-wide risk to develop the disease. Although PRS effectively capture  
588 established IBD variants, these scores may not account for yet unidentified genetic factors and  
589 were mainly derived from cohorts of predominantly European ancestry, thus missing out  
590 population specific variants [24]. However, in this cohort genetic risk for general IBD and CD  
591 showed stable and repeated connections to microbial- and physiological indicators of IBD and  
592 were derived and applied in a central European cohort. The relatively weak association of  
593 UC-PRS with microbial community characteristics, in contrast points to a reduced importance  
594 of the host-microbe interactions and/or lower density and penetrance of associated risk  
595 variants in UC [72]. In alignment with this, UC patients displayed comparatively weak and  
596 fewer associations with the microbiome, as seen before [64]. However, the associations that  
597 were detected are mostly consistent with other studies [66,73]. Naturally, PRS for IBDs are  
598 higher in IBD cases than in controls, however the clear and significantly elevated risk levels  
599 and the weak, albeit significant, increase in the probability of developing IBD with a higher  
600 PRS emphasizes the potential of these scores to stratify individuals at risk and underscores the  
601 close interaction between genetic risk and potential environmental triggers such as the  
602 microbiome in IBD development [17]. *Barnesiella sp.*, a common mucosa-associated butyrate  
603 producer, showed a positive association with CD risk in healthy individuals but was



604 negatively associated with CD risk in diseased individuals. A recent study revealed that this  
605 taxon is associated with a TLR4 modulating variant (*biliverdin reductase A, BLVRA*), which  
606 increases the risk for CD, but may dampen inflammation in a diseased state through SCFA  
607 production [74–76]. In contrast, we detected a reversal of this pattern in the commonly  
608 “probiotic” *Bacteroides fragillis* [77], which displayed a positive association with CD-PRS in  
609 IBD and CD patients. PRS also encompass important IBD-associated genes/risk variants for  
610 autophagy (e.g. *ATG16L1, NOD2*), which are unable to sense the anti-inflammatory signals of  
611 this taxon and may thus result in a proinflammatory cytokine bias through this taxon [78].

612 Broad genetic patterns, as determined by kinship, influenced the taxonomic and functional  
613 composition of the bacterial communities [79]. Social distance as well as broad relatedness  
614 have been described in detail to allow for bacterial transfer [80,81] and the transfer of  
615 potential phenotypes. Thus, horizontal and vertical intrafamily transmission [80], may add to  
616 the continuation and transfer of inflammation and dysbiosis as experimentally trackable in  
617 mouse experiments [82,83]. Interestingly the ubiquitous human gut commensal *Bacteroides*  
618 *uniformis* [57] among other *Bacteroides* species shows a strong association with host genetic  
619 variation as has been described elsewhere for humans and mice [84,76]. Although mainly  
620 commensals, which are associated to long chain carbohydrate degradation, this capability of  
621 *Bacteroides* may also drive pathogenic patterns in susceptible environments/hosts, due to the  
622 weakening of the mucosal barrier [85,86]. Thus, the varying genetic background of the host  
623 and strain/species specific functional variation may explain the varying association of  
624 *Bacteroides* species to IBD, particularly in light of their relatively stable host-genetic  
625 associations [76]. However, in contrast to our study, the heritability of Bacteroidetes in  
626 general and *Bacteroides* in particular were shown to be low in twin studies [79], but were on a  
627 similar level with recent studies which used more complex kinships [87]. Another mucolytic  
628 species, *Akkermansia muciniphila*, was highly heritable in other studies, as well [79,87], and

629 has often been associated with weight loss and metabolic health. Here we showed a clear  
630 association with non-IBD individuals, even though its role in IBD can be ambiguous [88,89].  
631 The majority of heritable taxa belong to Firmicutes (149 ASVs vs. 96 ASVs other); however,  
632 highly heritable *Christensenellaceae* were not detected, potentially due to their low  
633 abundance/prevalence in the cohort [81,79,90]. Overall, the potential inheritance patterns we  
634 detected strongly resemble patterns of shared communities in mother-offspring comparisons,  
635 thus potentially highlighting the dominant matrilineal influence of these analyses [80]],  
636 although cohabitation could not be completely disentangled from family membership due to  
637 data restrictions. In addition we detected a general correspondence between differences of  
638 community composition and genetic distance, highlighting a community wide influence of  
639 genetics and heritability. This pattern appears to be disturbed among IBD patients, potentially  
640 due to the overwhelming effect of disease and related factors on the microbial communities  
641 which masks the broad genetic effects we detected.

642 ***Oralization of the intestinal community and blooming of Enterobacteriaceae are common***  
643 ***IBD characteristics across cohorts***

644 A strong microbial pattern in IBD, particularly for CD, arises from the highly consistent  
645 association of oral bacteria with the disease. Various taxa across different phyla with a typical  
646 oral origin show clear associations with IBD (*Cand. Saccharibacteria*, *Rothia*, *Veillonella*,  
647 *Dialister*, *Fusobacteria*, *Klebsiella*) and highlighted a potential loss of colonization resistance  
648 of the community or detrimental physiological alterations. *Cand. Saccharibacteria* (formerly  
649 TM7) is a rarely described obligate bacterial epibiont involved in IBD-related microbial  
650 dysbiosis and oralization [37], but appeared very consistently in the investigated cohorts [64].  
651 Additionally, *R. muciliginosa* is an oral taxon with a clear association with IBD, which  
652 interestingly closely associates with *Cand. Saccharibacteria* (Figure S29). It may thus  
653 represent a potential host for this taxon in the intestinal environment, as it until recently

654 belonged to the same family as its oral hosts (*Actinomycetaceae*) [91]. Alternatively, this  
655 group may represent an independent TM7 group adapted to the intestinal environment, as  
656 observed in the ruminant TM7-G3 cluster [92] with its own host species. *F. nucleatum* and its  
657 cell components have been proposed as causal triggers of inflammation in e.g. rheumatoid  
658 arthritis [93] or even IBD [94], and have been implicated in the development of colorectal  
659 cancer [95]. Recent studies also displayed several overlapping, mostly oral bacteria associated  
660 to UC, such as *Veillonella dispar*, *Megasphaera micronuciformis*, *Haemophilus*  
661 *parainfluenzae*, and other *Enterobacteriaceae* [96] and a reduction of these ectopically  
662 colonizing oral bacteria can promote remission in IBD patients [66]. Furthermore, oralization  
663 of the microbiome and oral health itself are strongly associated to IBD development, risk, and  
664 severity [97,98].

665 However, the most consistent pattern of IBD association was observed for  
666 *Enterobacteriaceae*, such as potentially oral *Klebsiella sp.*, which are highly proinflammatory  
667 and often associated with UC and PSC, although we mainly detected stable associations with  
668 CD and IBD [97]. In addition, *Escherichia/Shigella* show clear and valid pathogenic or  
669 proinflammatory characteristics [38,99] and appear to characterize a distinct, potentially  
670 dysbiotic community type (DMM cluster-3), as previously proposed [44,45]. Various  
671 *Enterobacteriaceae*, are known to elicit strong immunological reactions and are associated  
672 with the development of IBD at various levels ranging from ecological (facilitation [100]) to  
673 metabolic [101] interactions that allow these taxa to benefit from inflammatory processes  
674 [102]. However, some members of the *Escherichia/Shigella* have been shown to be beneficial  
675 for the immunological development early in life [103] or carry beneficial characteristics  
676 [104]. Differentially abundant taxa, particularly oral taxa (*Fusobacterium*, *Dialister*,  
677 *Veillonella*) and the abundance patterns of *Enterobacteriaceae* correspond well with a recent

678 large meta-analysis of IBD studies, as do abundances of potentially beneficial taxa (e.g.  
679 *Faecalibacterium*, *Ruminococcaceae*, *Alistipes*) [105].

680 ***Amino acid metabolism and flagellar functions associate strongly with healthy individuals***

681 Interestingly, one of the most enriched functions differing between healthy and diseased  
682 individuals is *Aminobenzoate Metabolism*. Anthranilate or 2-Aminobenzoic acid is a key  
683 metabolite of bacterial aromatic compound degradation, a precursor of tryptophan  
684 biosynthesis, and a key metabolite to reduce or even prevent IBD pathogenesis [106,107].  
685 Additionally, other functions involved in amino acid- and tryptophan metabolism, e.g.  
686 *Phenylalanine*, *tyrosine* and *tryptophan biosynthesis*, were also enriched. Thus, aromatic  
687 amino acid biosynthesis appears to have a strong, and health relevant effect, as the enrichment  
688 of amino acid metabolism is very prominent in control subjects. Interestingly, bacterial  
689 functions involved in *Flagellar assembly* show a strong and consistent association with  
690 healthy subjects as well (antagonistic to IBD) and may arise from the abundance of so-called  
691 “silent” flagellins present in *Lachnospiraceae*, which have potential anti-inflammatory  
692 activity via TLR5 binding [108]. However, *Lachnospiraceae* derived flagellins display a  
693 strong immune reactivity in CD patients [109,110], which may represent an aberrant response  
694 driven by an already disturbed mucosal barrier, rather than a *de novo* reaction to members of  
695 the *Lachnospiraceae* [111,109]. *Clostridium clostridioforme* (*Clostridium XIVa*) is highly  
696 CD-specific and immunologically reactive according to our results and other studies  
697 [112,113]. A recent study identified a strong and cross cohort IBD association of  
698 *C. clostridioforme* with specific biosynthetic gene clusters, whose products lead to increased  
699 inflammation and mucosal barrier dysfunction [114]. *C. clostridioforme* strongly correlates  
700 with the concentration of the potentially pro-inflammatory long chained fatty acid  
701 eicosatetraenoic acid/arachidonic acid (F20:4) [113], and its uptake is positively associated to  
702 dysbiosis (e.g. long saturated: long unsaturated: lipoids) correlated with community

703 composition and dysbiosis. Furthermore eicosatetraenoic acid positively associated with CD-  
704 and UC-associated taxa, *R. gnavus* but also *C. clostridioforme*, whereas typical probiotic taxa  
705 showed a significant negative relationship with this metabolite [113]. In addition to its  
706 immune- and inflammatory characteristics, this fatty acid possesses bactericidal activity and  
707 the potential to disrupt bacterial cell membranes [115]. This might indicate that IBD-  
708 associated taxa might be able to metabolize and synthesize F20:4 and are less susceptible to it,  
709 while it interferes with the growth health-associated species [113,114]. In contrast, the  
710 consumption of vitamins, long-chain carbohydrates, and long unsaturated fatty acids is  
711 correlated with lower levels of dysbiosis [42,116]. Bacterial traits that are central to the  
712 generation and integration of immune modulatory factors of gram-negative  
713 (*Lipopolysaccharide biosynthesis*) and gram-positive bacteria (*Teichoic acid biosynthesis*,  
714 *Peptidoglycan biosynthesis*) on the bacterial surface are also consistently enriched among IBD  
715 patients. Comparable patterns of these surface/capsular and membrane factors have also been  
716 observed in a recent meta-analysis of various IBD cohorts [117].

717

## 718 CONCLUSION

719 The KINDRED cohort is a prospective cohort study that collects biomaterial and various  
720 types of data from IBD patients and their first- or second-degree relatives. The main aim of  
721 the KINDRED cohort is to identify lifestyle factors as well as serum and stool markers  
722 associated with the onset of the disease (in initially healthy relatives with a positive family  
723 history for IBD) and to investigate the individual disease course in patients on a multiomic  
724 level, integrating lifestyle, genetics, and microbial information. Using this multiomic dataset  
725 we identified strong and generalizable dysbiosis gradients, which correspond strongly with  
726 IBD pathologies, physiological manifestations of inflammation (*e.g.* BSS, calprotectin, anti-  
727 GP2 and ASCA IgA/IgG), genetic risk for IBD and general risk of disease onset. Patterns of

728 overabundance and importance of various opportunistic pathogens (*e.g.*, *Enterobacteriaceae*,  
729 *C. XIVa clostridioforme*), in addition to consistent patterns of oralization, characterize IBD  
730 patients. Functionally, pathways involved in amino acid metabolism and flagellar assembly  
731 are beneficial, while mucolytic functions are associated with IBD. Broad scale ecological  
732 patterns point totipping-point dynamics being involved in the drastic state transitions of  
733 communities into the chaotic communities characteristic for IBD (community variability,  
734 community differentiation). However, some limitations are present in this study. Healthy  
735 relatives were not examined in an examination center and all information about these  
736 individuals were obtained through self-administered questionnaires. Another potential  
737 weakness of the KINDRED cohort is its open design, where recruitment has run over a long  
738 period of time (from 2013, and ongoing), with a slow increase of onset cases, and steady  
739 decrease of participants with time due to lack of compliance or withdrawal from the study  
740 during the regular, yet only biennial follow-ups.

741

## 742 **MATERIALS AND METHODS**

743 **Basic study design and cohort:** The KINDRED cohort is a German-wide, prospective cohort  
744 study that collected both, questionnaire data and biomaterials from IBD patients and their  
745 (affected and unaffected) family members. Despite great efforts to include as many family  
746 members as possible, a focus has always been on unaffected individuals with a family history  
747 of IBD. The recruitment of study participants started in October 2013 and is still ongoing.  
748 Follow-up information and new biomaterial samples were collected prospectively at intervals  
749 of approximately two years (Figure S1-S3). As of April 2021, the IBD Family Cohort has  
750 thus enrolled 1497 IBD patients together with 1813 (initially) non-affected first- or second-  
751 degree relatives from Germany (minimum age at inclusion: 7 years). Participants, including  
752 the IBD patients, were asked to provide questionnaire data and biomaterials (blood, stool,

753 hair) at baseline and after every 2 years of follow-up (Table S1 & S2). The resulting patient  
754 counts are shown in Table 1 and Figure 1A. In addition, physician-administered  
755 questionnaires were collected from IBD patients to obtain physician-validated information,  
756 such as diagnosis, disease pattern/location, activity, and medication (Table S1). Using these  
757 data, the Kiel IBD Family Cohort aims to facilitate the comprehensive molecular, clinical,  
758 lifestyle, nutritional, and sociodemographic characterization of patients and high-risk  
759 individuals, and to identify preclinical signs for the onset of IBD. Additional information on  
760 eligibility criteria, enrollment, data and biomaterial collection, as well as data management,  
761 and privacy protection is given in the Supplement of this article. A standard battery of  
762 questionnaires was used to assess self-reported dietary behavior (12 month recall  
763 questionnaire) and metrics of well-being and quality of life (quality of life index, Fatigue  
764 Severity Scale index), were also employed (see Supplement) [118–120]. Nutritional data were  
765 adjusted by total energy consumption to decouple caloric consumption from diet composition.  
766 Individuals with undefined or unclear IBD diagnosis but with intestinal inflammation (e.g.  
767 suspected CD/UC, colitis) are summarized under the category uIBD (undefined IBD, Table  
768 S3).

769 **Ethics and human samples:** The KINDRED cohort study protocol was approved by the  
770 ethics committee of the Medical Faculty of Kiel University (AZ A117/13). Every study  
771 participant provided written informed consent on forms that were age-adapted. For  
772 participants under the age of 18 years, informed consent must also be signed by their parents.  
773 When the participants reached the ages of 12 and 18 years a new informed consent  
774 (reconsent) form was signed by the participants themselves and, in the case of 12-year-old  
775 adolescents, by their parents.

776 A cohort of treatment naïve and newly diagnosed IBD patients in an active disease state from  
777 Malta (naïve :  $N_{CD}=31$ ,  $N_{UC}=25$ ), including a healthy control cohort ( $N_{Contr.}=96$ ), was recruited

778 as described elsewhere [64]. In addition we included a Maltese patient cohort currently in  
779 disease remission and treatment (remission:  $N_{CD}=32$ ,  $N_{UC}=66$ ), as described recently [63].

780 Individuals from the Swedish Inception Cohort in IBD (SIC IBD) were included as treatment-  
781 naïve patients, between 20–77 years of age. Symptoms, such as diarrhea, abdominal pain,  
782 bloody or mucous stools for >2 weeks, were inclusion criteria. The final diagnosis of IBD was  
783 established according to internationally accepted criteria including clinical, microbiological,  
784 endoscopic, histological, and radiological evaluation ( $N_{CD}=17$ ,  $N_{UC}=16$ ). Patients with  
785 gastrointestinal symptoms but without endoscopic and histological signs of IBD-associated  
786 inflammation were considered symptomatic controls (SC,  $N_{SC}=16$ ). In addition, 17 healthy  
787 individuals were included as healthy controls ( $N_{contr.}=16$ , one failed sequencing).

788 **Physiological measurements:** Calprotectin (indicator of intestinal inflammation) was  
789 measured in fecal samples using a Bühlmann fCAL™ ELISA kit (BÜHLMANN  
790 LABORATORIES AG) and analyzed using SoftMax Pro Software (Molecular Devices).  
791 ASCA IgA/IgG and anti-GP2 IgA/IgG were measured externally using ELISA (Medipan  
792 GmbH) including batchwise calibration samples. Occult fecal blood was determined by  
793 PreventID® Haemo/HaptOccult (Preventis GmbH, Bensheim, Germany). Other  
794 immunological measures (CRP Hb) were obtained during the initial examination using  
795 standard clinical tests.

796 **Polygenic Risk Scores:** GSA data were quality controlled via gwas-qc  
797 (<https://github.com/ikmb/gwas-qc>) not correcting for closely related individuals in the cohort  
798 using hg19 (genome build 37) of the human genome and 1000 Genome reference set.  
799 Potential sample mix ups and unclear relatedness patterns were manually checked and  
800 corrected if needed. Imputation was performed after chromosome wise transformation into  
801 bgzip VCF files via plink2 [121]. Single VCF files were uploaded to  
802 [hybridcomputing.ikmb.uni-kiel.de/webservice/sites/](https://hybridcomputing.ikmb.uni-kiel.de/webservice/sites/), imputed and phased via EagleImp



803 (Genome build: GRCh37/hg19, Reference: 1000 Genomes Phase 3,  $r^2$  filter: 0.1, allowed  
804 reference swaps, and strand flips) [122].

805 To examine the genetic susceptibility of individuals in the KINDRED cohort for IBD and its  
806 subtypes CD and UC we calculated Polygenic Risk Scores (PRS). GWAS summary statistics  
807 were taken from the meta-analysis of IBD from Liu et al [20]. The summary statistics are  
808 based on a total of 5956 CD and 6968 UC patients with an additional 21770 population of  
809 controls with European ancestry. PRS were calculated with *LDpred2* within the *bigsnpr* (v.  
810 1.12.2) R package [123,36]. After quality control of the summary statistics [124] the method  
811 calculates a posterior mean effect size based on linkage disequilibrium information and base  
812 effect size for all remaining available markers in both data sets. We used the *auto*-method of  
813 *LDpred2*, which automatically estimated the parameter sparsity  $p$  and the SNP heritability  $h^2$   
814 and did not require to tune hyper-parameters in the validation set.

815 **Stool sample processing:** DNA extraction, sequencing and bioinformatics processing of 16S  
816 rRNA gene libraries from stool samples were performed as described previously in detail  
817 [125].

818 **Data processing:** Data processing of 16S sequences was performed using DADA2 1.10 [126]  
819 via the workflow for big datasets (<https://benjjneb.github.io/dada2/bigdata.html>,  
820 [https://github.com/mruehlemann/ikmb\\_amplicon\\_processing/blob/master/  
821 dada2\\_16S\\_workow\\_with\\_AR.R](https://github.com/mruehlemann/ikmb_amplicon_processing/blob/master/dada2_16S_workow_with_AR.R)) resulting in abundance tables of Amplicon Sequence  
822 Variants (ASVs). All sequencing runs were handled separately for error correction, read  
823 merging, and combined chimera detection. ASVs underwent taxonomic annotation using the  
824 naïve Bayesian classifier implemented in DADA2 using the Ribosomal Database Project 16  
825 release [127,128]. ASV sequences were aligned via NAST-alignment to the SILVA core  
826 database and filtered for informative sites (constant gaps, constant bases) in mothur [129].  
827 Phylogenetic tree construction on ASV alignment generated was carried out using FastTree

828 2.1 with the CAT substitution model with  $\Gamma$ -correction and improved accuracy, employing  
829 more minimum evolution rounds for initial tree search [-spr 4], more exhaustive tree search [-  
830 mlacc 2], and a slower initial tree search [-slownni] [130]. Unstratified KO categories of  
831 metagenome predictions were generated using the native ASV abundances and sequences in  
832 PICRUST 2.5.0 using default workflows [131].

### 833 **Statistical methods:**

834 **Alpha diversity:** Species richness (Chao1), Simpson diversity (1-D), and phylogenetic alpha  
835 diversity (Nearest Taxon Index [NTI], Net Relatedness Index [NRI]) were calculated and  
836 analyzed in R 3.5.3.17-19. Phylogenetic measures of alpha diversity [Nearest Taxon Index  
837 (NTI), and Net Relatedness Index (NRI)] were derived using the *PhyloMeasures* package,  
838 based on 999 permutations against a null model preserving relative species richness within the  
839 communities [132–134] [ $NRI = -1 \times (MPD_{observed} - \text{mean}(MPD_{random})) / SD(MPD_{random})$ ];  
840  $NTI = -1 \times (MNTD_{observed} - \text{mean}(MNTD_{random})) / SD(MNTD_{random})$ ]. Both metrics are  
841 phylogenetic effect sizes, for which positive values indicate phylogenetic clustering, values  
842 close to zero indicate neutral or random community assembly, and negative values indicate  
843 phylogenetic overdispersion, either over the whole phylogenetic tree (NRI) or across the  
844 closest related species/tips of the phylogenetic tree (NTI). The relationships of physiological-  
845 and microbial inflammation markers/indices with alpha diversity were analyzed using linear  
846 models after correcting for relevant covariates (LM: residuals(variable~gender+scaled BMI +  
847 scaled Age) ~ IBD (Contr., CD, UC, IBD) \* physiological variable). Model fits were  
848 visualized via *base* R and *jTools* [135]. General correlation of MD with alpha diversity was  
849 tested via simple linear models, segmented linear models (*segmented*), or polynomial linear  
850 models and assessed via the AIC.

851 **Beta diversity:** Analyses were conducted via distance based (conditional) Redundancy  
852 analyses and permutative ANOVA, as well as with a multivariate test for homogeneity of

853 variances (10000 permutations) [136,137] using Bray-Curtis dissimilarity (differential  
854 abundance). Global and pairwise differences in community variability were assessed via  
855 permutation test of multivariate homogeneity of group dispersions (10'000 permutations, via  
856 the *betadisper* function). Community clustering according to anthropometric, community  
857 distances of naïve and rarefied communities were highly correlated and thus higher coverage  
858 naïve samples were used for analysis (Bray-Curtis: Mantel:  $r=0.9588$ ,  $P<0.001$ ; Procrustes:  
859  $m^2=0.06012$ ,  $r=0.9695$ ,  $P<0.001$ ; Jaccard: Mantel:  $r=0.9941$ ,  $P<0.001$ ; Procrustes:  
860  $m^2=0.006143$ ,  $r=0.9969$ ,  $P<0.001$ ).

861 **DMM community clustering:** Community clustering was performed via Dirichlet  
862 Multinomial Mixture modeling as implemented in the R packages *microbiome*  
863 (*microbiome\_1.20.0*) [138] and *DirichletMultinomial* (*DirichletMultinomial\_1.40.0*) [139]  
864 after centered log ratio transformation (CLR) implemented in *compositions* [140]. The  
865 optimal number of clusters was determined via Laplace goodness of fit optimization in a  
866 range of 1-15 clusters.

867 **Differential abundance analysis:** Taxon abundances were filtered by having normalized  
868 counts of at least 5 among 1% of samples (*DESeq2* median of ratios transformation). Negative  
869 binomial GLMs with Wald tests as implemented in *DESeq2* (including automated effect  
870 filtering) were employed to detect differentially abundant taxa for each time point separately  
871 as well as for all time points combined [141]. To reduce the effect of potential confounding  
872 effects we included subject age, BMI and sex as covariates and in addition the time point  
873 itself, for the combined dataset. P-values were adjusted via FDR and only comparisons  
874 excluding the small subset of uIBD patients were investigated further (Contr./CD, Contr./UC,  
875 CD/UC, Contr./all-IBD).

876 Abundance~sex+ BMI (centered-scaled) + age (centered-scaled) [+ time point (BL/F1/F2)] +  
877 IBD pathology (Control, CD, UC, uIBD)

878 Abundance~sex+ BMI (centered-scaled) + age (centered-scaled) [+ time point (BL/F1/F2)] +  
879 Health status (Control/[CD, UC, uIBD])

880 Similar methods were applied to investigate the functional characteristics of the bacterial  
881 communities, as based on imputed bacterial functions and pathways via PICRUSt2 [131]. To  
882 generalize the single KO associations, we performed Gene Set Enrichment Analyses using a  
883 function wise score ( $-\log_{10}(P_{\text{Wald-test}}) \times \text{sign}(\log \text{ fold change})$ ) by multiplying the natural  
884 logarithm of P-values multiplied by the sign of fold change via clusterProfiler (4.4.4) using  
885 the *gseKEGG* function [47,142].

886 Taxon abundances were correlated with a reduced set of clinical variables (ASCA IgA, ASCA  
887 IgG, anti-GP2 IgA, anti-GP2 IgG, fecal calprotectin, BSS) as well as polygenic risk scores  
888 for IBD in general and for CD or UC in particular. CLR transformed species abundances were  
889 correlated with core physiological measures via *ppcor* [143], combining the P-values of  
890 Spearman-, Kendall-, and Pearson correlations via Brown's method to detect and incorporate  
891 the most consistent associations across different association measures [144]. The resulting P-  
892 values were FDR corrected and Spearman  $\rho$  was used for visualization via *ComplexHeatmap*  
893 (v2.14.0) [145].

894 **Dysbiosis score:** Dysbiosis scores based on the distribution of bacterial community members  
895 were calculated following Gevers *et al.* (2014) (Microbial Dysbiosis index/MD-index) and the  
896 more generalized General Microbiome Health Index (GMHI) as developed by Gupta *et al.*  
897 (2020) [38,39]. For the MD-index we combined significant taxa abundant in either CD or UC  
898 patients (uIBD not included) as compared to healthy/non-IBD cohort members at the baseline  
899 time point of the KINDRED cohort [ $\log_{1p}(\text{sums}(\text{abundance CD}, \text{abundance UC})) /$   
900  $\log_{1p}(\text{sums}(\text{non-IBD}))$ ]. GMHI was trained on the relative abundances at the baseline time  
901 point (training: 90%, testing: 10%) of the KINDRED cohort using code adapted from the

902 Dutch Microbiome Project

903 (["https://github.com/GRONINGEN-MICROBIOME-CENTRE/DMP/tree/main/gmhi"](https://github.com/GRONINGEN-MICROBIOME-CENTRE/DMP/tree/main/gmhi)).

904 Physiological and microbial inflammation markers/indices were analysed using linear models  
905 including relevant covariates (variable~gender+scaled BMI + scaled Age + IBD (Contr., CD,  
906 UC, IBD)), followed by stepwise model selection to minimize AIC without significant loss of  
907 fit. The correlation of the MD/GMHI with alpha diversity was assessed with linear and  
908 quadratic fits and the best fit was selected via minimal AIC for each time point. Model fits  
909 were visualized via *base* R and *jTools* [135].

910 **Heritability analyses:** To estimate the heritability of single communities in a rather  
911 heterogeneous and patchy pedigree with many confounding variables we applied a recently  
912 published approach employing linear mixed models, as implemented in *lme4qtl* (*relatLmer*)  
913 [146]. Thus, we were able to incorporate fixed effects in our heritability estimates using a  
914 kinship matrix based on reported pedigree information (*kinship2*) (Sinnwell et al., 2014) and  
915 CLR transformed taxon abundances. Evaluating the models with and without covariates (AIC,  
916 based on ML fit) allows us to further evaluate whether taxon heritability/family association is  
917 influenced significantly by environmental factors or family relations alone. To assess patterns  
918 among close relatives we categorized relatedness patterns (related (>1st degree), first degree  
919 (sibling, parent)) for comparison, as well as calculated distances between healthy family  
920 members and diseased and eventually diseased individuals (onsets). Similar techniques were  
921 applied for the analyses of KO heritability, including subject age and sex as covariates. Gene  
922 Set Enrichment Analyses were performed on K-numbers using a function wise score based on  
923 full  $h^2$  estimates via clusterProfiler (4.4.4) using the *gseKEGG* function [142]. Community-  
924 wide correspondence between taxonomic and functional community dissimilarity and  
925 relatedness (kinship matrix, plink2 based "Identity by State" distance derived from imputed  
926 SNP sets) was assessed via PROCRUSTES tests (5000 permutations).

927 **Network analyses:** Networks were generated individually for healthy and IBD (UC, CD)  
928 cohort subsets, as well as for the complete time point subsets. The node-based values (degree,  
929 betweenness, PageRank-index, eigenvalue-centrality, k-nearest neighbor degree) were  
930 calculated in *igraph* 1.2.4.1 [147–149]. Network-wide measures include centrality based  
931 assortativity [150], network diameter, radius, and size, density/clustering [151,152], and  
932 natural connectivity [48]. To assess whether bacteria were more important than expected by  
933 chance, observed centralities (mean of control subsamples) were compared against a  
934 permuted set of networks (10,000 times, combined for control subsamples) via one-sided Z-  
935 tests. Graphlet (4-node) frequency correlation (*orca* v1.1-1, Spearman) based Euclidean  
936 distance [153,154] and an edge sharing distance (frequency of shared pairwise correlations) to  
937 assess network similarity between the different subgroup networks (CD, UC, Contr., IBD at  
938 baseline, follow-up 1, follow-up 2).

939 **Onset association:** To predict the binary outcomes of disease onset based on diversity  
940 measures and physiological/clinical variables, we employed generalized linear models with a  
941 binomial error structure and a “clog-log” link function (“complementary-log-log”) as  
942 implemented in MASS [155]. The nonsymmetric complementary log-log link function (“clog-  
943 log”) was chosen, for its ability to better cope with unbalanced distributions between positive  
944 and negative outcomes in the target variable [156]. Baseline samples with the addition of one  
945 follow-up 1 sample, which developed disease in the second follow-up, were used for  
946 prediction ( $N_{\text{onset}}=7$ ), in the context of all remaining baseline samples or only healthy baseline  
947 controls. Models included either no covariates, or the established set of covariates (scaled age,  
948 sex, scaled BMI). Model fits were visualized via *base* R and *jTools* [135].

949 Models:

950 IBD-onset next time point (1/0)~ metric

951 IBD-onset next time point (1/0)~sex+ BMI (centered-scaled) + age (centered-scaled) +  
952 metric/clinical variable

953 Data structure: BL ( $N_{\text{onset}}=4$ ,  $N_{\text{CD}}=551$ ,  $N_{\text{UC}}=438$ ,  $N_{\text{ulIBD}}=32$ ,  $N_{\text{contr.}}=787$ ); BL-controls ( $N_{\text{onset}}=4$ ,  
954  $N_{\text{contr.}}=787$ ); BL+ ( $N_{\text{onset}}=7$ ,  $N_{\text{CD}}=551$ ,  $N_{\text{UC}}=438$ ,  $N_{\text{ulIBD}}=32$ ,  $N_{\text{contr.}}=787$ ); BL-controls+ ( $N_{\text{onset}}=7$ ,  
955  $N_{\text{contr.}}=787$ )

956

## 957 **ACKNOWLEDGEMENTS**

958 We would like to thank Ilona Urbach, Ines Wulf, and Tonio Hauptmann of the IKMB  
959 microbiome laboratory and the staff of the IKMB sequencing facilities for their excellent  
960 technical support. This study was supported by the Deutsche Forschungsgemeinschaft (DFG)  
961 Research Unit 5042 (“miTarget-The Microbiome as a Therapeutic Target in Inflammatory  
962 Bowel Diseases”), the University of Malta, and the University of Örebro. Special thanks go to  
963 all participants of the KINDRED cohort study for enabling such a unique prospective study.

964

## 965 **DECLARATIONS:**

966 **Funding:** The KINDRED cohort received funding from the German Research Foundation  
967 (DFG) through Excellence Clusters “Inflammation at interfaces” (EXC 306) and “Precision  
968 Medicine in Inflammation” (PMI; EXC 2167), as well as from the Research Unit miTarget  
969 (FOR 5042). This study was also supported by a grant from the German Ministry of  
970 Education and Research (01ZX1606A).

971 **Conflicts of interest:** The authors declare that they have no conflicts of interest related to this  
972 article.

973 **Availability of data and material:** All KINDRED cohort data are stored and managed by the  
974 PopGen biobank at the Institute of Epidemiology at Kiel University, Germany. Researchers  
975 can apply for data access to the KINDRED data by submitting a research proposal, including  
976 the scientific background, research question, success prospects, study design, potential  
977 conclusions, and scientific collaborators of their study, at the local biobank P2N via the  
978 following form: <http://www.uksh.de/p2n/Information+for+Researchers.html>. Due to the  
979 informed consent obtained from the participants, phenotypes, as well as genotyping and all  
980 16S rRNA gene-sequencing data, can not be deposited publicly.

981 Raw sequence data and relevant meta-data can be accessed online under the accession number  
982 PRJEB44440 (Malta IBD cases in remission), PRJEB47161 (Malta treatment naive IBD  
983 cases), PRJEB47162 (Malta controls) [63,64]; and data of the Swedish SIC-IBD inception  
984 cohort will be made available under the accession number PRJEB77933 at the European  
985 Nucleotide Archive (<https://www.ebi.ac.uk/ena/>) after acceptance of this manuscript at a peer-  
986 reviewed journal.

987 **Ethics approval:** The study was performed in accordance with the principles of the  
988 Declaration of Helsinki. Approval was granted by the ethics committee of the Medical  
989 Faculty of Kiel University (AZ A117/13). Ethical approval for the Maltese cohorts was  
990 obtained from the University of Malta Research Ethics Committee (Ref 32/2017) and patients  
991 were recruited from the gastroenterology outpatient clinic at Mater Dei Hospital, Malta  
992 [63,64]. Ethical approval for the Swedish SIC-IBD cohort was obtained from the Uppsala  
993 Regional Ethics Committee (2010/313) and suspected IBD patients and controls were  
994 recruited from gastroenterological units at six Swedish hospitals.

995 **Consent to participate:** Informed consent was obtained from all individual participants  
996 included in the study (or, in the case of children, from their parents). Consent for publication:  
997 Not applicable.



998 **Authors' contributions:** GJ, SS, MK, CB, AF and WL designed the study; JE, GJ, and WL,  
999 LT and KJ conducted the study; LT, KJ and PRa curated the data; PE, RK, JJ and DR  
1000 contributed external samples and additional clinical analyses; PRa and IR analyzed the data;  
1001 PRa, IR, CB, WL and AF drafted the paper; all authors take responsibility for the final  
1002 content. All authors have read the manuscript, made significant intellectual contributions and  
1003 approved the final version of the manuscript.

1004

## 1005 FIGURES

1006 **Figure 1: (A)** Individual transitions in the cohort between time points. The highlighted  
1007 transitions are onset cases detected during the study's runtime ( $N_{BL \rightarrow F1}=4$ ,  $N_{F1 \rightarrow F2}=1$ ,  
1008  $N_{F2 \rightarrow F3}=2$ ,  $N_{F3 \rightarrow F4}=3$ ). **(B)** Analysis of selected physiological inflammation markers (ASCA  
1009 IgA/IgG, Calprotectin, BSS) with respect to IBD condition and relevant covariates via linear  
1010 models at baseline (see Table 1). Shown are only the optimal model results, after the model  
1011 selection procedures are shown. Pairwise differences with respect to average differences, or  
1012 differences in slope are highlighted in the plots (#  $P \leq 0.1000$ , \*  $P \leq 0.0500$ , \*\*  $P \leq 0.0100$ ,  
1013 \*\*\*  $P \leq 0.0010$ ). **(C)** The violin plot displays the average differences of genetic predisposition  
1014 to CD, UC and IBD in general, as based on *LDpred2* derived PRS. The average differences in  
1015 PRS were tested via Wilcoxon rank tests contrasting healthy individuals, with healthy future  
1016 onset cases (grey), CD patients (red), UC patients, and patients with undefined IBD (uIBD).  
1017 In particular, compared with those in healthy controls, the average risk of IBD is significantly  
1018 and consistently greater in patients with CD and UC, as well as in patients with future-onset  
1019 disease. **(D)** Scatterplots show the significant relationships between polygenic risk scores for  
1020 CD (PRS-CD) and selected physiological inflammation markers, between PRS-CD and ASCA  
1021 IgG by considering anthropometric covariates (linear model on residuals) or without

1022 (Spearman correlation). Additional correlations of IBD- and UC-PRS with physiological  
1023 markers can be found in Figure S4 and Table S5.

1024

1025 **Figure 2:** Taxon differential abundances and predictability. **(A)** Overview of individual  
1026 phylum abundances across time points and health conditions in the KINDRED Cohort. **(B)**  
1027 Differential abundance analyses of ASVs based on the sample time point (BL). Displayed are  
1028 the log fold changes for each ASV clustered by genus classification, including standard errors  
1029 of the fold changes as indicated by individual error bars. Color indicates the phylum  
1030 membership of the ASVs/Genera. DA only displays significant differential abundance for the  
1031 respective comparison/contrast ( $P_{FDR} \leq 0.05$ ). **(C)** Partial correlation of CLR transformed taxon  
1032 abundances with core physiological measures via *ppcor* [143], combining the *P* values of  
1033 Spearman-, Kendall-, and Pearson correlations via Brown's method and corrected via FDR  
1034 [144]. Correlations were adjusted for age, gender, and BMI and Spearman  $\rho$  is used to  
1035 visualize correlation strength between taxa and clinical measures ( $\# P_{FDR} \leq 0.1000$ ,  
1036  $* P_{FDR} \leq 0.0500$ ,  $** P_{FDR} \leq 0.0100$ ,  $*** P_{FDR} \leq 0.0010$ ). Overlapping patterns of differential  
1037 abundance for the respective taxa in the KINDRED cohort, Maltese-, and Swedish cohort is  
1038 indicated in the bottom color bars. **(D)** Differences in species richness between healthy  
1039 individuals, future onset cases, CD, UC, and uIBD cases (Wilcoxon test). Correlation of  
1040 species richness with the microbial dysbiosis index (MD-index), which results in a negative,  
1041 but non-linear relationship (optimal AIC based fit) between diversity and dysbiosis (BL  
1042 (poly):  $F_{2,1809} = 226.97$ ,  $P < 2.2 \times 10^{-16}$ ,  $adj.R^2 = 0.1997$ ; F1 (poly):  $F_{2,645} = 83.778$ ,  $P < 2.2 \times 10^{-16}$ ,  
1043  $adj.R^2 = 0.2037$ ; F2 (poly):  $F_{2,536} = 55.745$ ,  $P < 2.2 \times 10^{-16}$ ,  $adj.R^2 = 0.1691$ ; LM). Also the  
1044 relationship between alpha diversity and several IBD relevant clinical measures, shows  
1045 significant relationships between alpha diversity and host physiology, in a disease condition  
1046 specific manner (see Table S15, Figure S11). Pairwise differences with respect to average

1047 differences, or differences in slope are highlighted in the plots (#  $P \leq 0.1000$ , \*  $P \leq 0.0500$ ,  
1048 \*\*  $P \leq 0.0100$ , \*\*\*  $P \leq 0.0010$ ).

1049

1050 **Figure 3: (A)** Non-metric Multidimensional Scaling (NMDS) of Bray-Curtis distances among  
1051 baseline samples, displaying the significant clustering by health conditions and significant  
1052 correlations of clinical inflammation measures with community distance (BL, see Table S16  
1053 & Table S17). **(B)** NMDS displaying the gradient of community dysbiosis as expressed by  
1054 MD-index [38], in parallel with clinical measures of inflammation and healthy onset cases  
1055 highlighted in red (\*, develop IBD in F1). **(C)** Correlation of MD-index and the first NMDS  
1056 axis showing a clear gradient of dysbiosis in the community. Onset cases are distributed in the  
1057 range of standard deviation around the mean of the community distribution (NMDS1) and the  
1058 severity of dysbiosis (MD-index). **(D)** Community variability between health conditions as  
1059 measured by the distance to the group centroid, is overall significantly different between  
1060 health conditions ( $F_{3,1808}=46.0315$ ,  $P=0.00001$ , PERMANOVA), and significantly increased in  
1061 CD and UC patients as compared to healthy controls (Table S19). **(E)** Principle coordinate  
1062 analysis of German-, Swedish-, and Maltese samples, highlighting the transferrability of the  
1063 dysbiosis gradient across cohorts (MD-index derived from german samples), **(F)** as well as a  
1064 common disease wise clustering of communities irrespective of sample origin. **(G)**  
1065 Community variability between health/IBD conditions within and between the German-,  
1066 Swedish-, and Maltese cohorts showing a common theme of increased variability in IBD  
1067 cases. **(H)** Mean differences of dysbiosis (MD-index derived from german samples) within  
1068 and across cohorts, with the strongest differences between healthy and CD individuals. **(I)**  
1069 Scatterplots visualize the correlation of selected physiological inflammation markers with the  
1070 microbial dysbiosis score [38], and show disease specific differences as compared to healthy  
1071 control individuals. Pairwise differences with respect to average differences, or differences in

1072 slope are highlighted in the plots (see Figure S16, Table S18; #  $P \leq 0.1000$ , \*  $P \leq 0.0500$ ,  
1073 \*\*  $P \leq 0.0100$ , \*\*\*  $P \leq 0.0010$ ). **(J)** Visualization of the explained variation of significant  
1074 anthropometric variables as based on serial PERMANOVA of Bray-Curtis distances in all  
1075 three time points available and focused on physiological measures (Table S17), different  
1076 reported pathologies of individuals (Table S21), use of pharmaceuticals (Table S22), and  
1077 nutrient intake as derived from FFQ data (Table S23). Variables are displayed if they show  
1078 significant clustering in at least one time point (#  $P_{FDR} \leq 0.1000$ , \*  $P_{FDR} \leq 0.0500$ ,  
1079 \*\*  $P_{FDR} \leq 0.0100$ , \*\*\*  $P_{FDR} \leq 0.0010$ ).

1080

1081 **Figure 4:** **(A)** Community clusters of the microbial community at the baseline time point,  
1082 determined by Dirichlet Multinomial Mixture modelling (DMM) and optimal clustering was  
1083 determined via Laplace goodness of fit optimization. **(B)** Overlay of the Microbial Dysbiosis  
1084 score gradient (Gevers *et al.*, 2014) and community clusters (outlines), including healthy  
1085 onset patients at baseline (indicated by \*). **(C)** Community clusters display a significantly  
1086 elevated level of dysbiosis in clusters 2 and 3 (pairwise Wilcoxon tests), **(D)** as well as  
1087 elevated levels of community variability in cluster-3 (PERMANOVA). **(E)** Non-metric  
1088 Multidimensional Scaling (NMDS) of Bray-Curtis distances of PICRUSt2 based KO  
1089 abundances among baseline samples (Douglas *et al.*, 2020), displaying the gradient of  
1090 community dysbiosis as expressed by the taxonomy based MD-index (Gevers *et al.*, 2014),  
1091 **(F)** and displaying the significant clustering by health conditions and significant correlations  
1092 of clinical inflammation measures with community distance (see Table S16 & Table S17).  
1093 Healthy IBD onset patients are highlighted in red (\*, patients who developed IBD by the next  
1094 follow-up). **(G)** Community variability between health conditions as measured by the distance  
1095 to the group centroid, was significantly increased in CD and UC patients as compared to  
1096 healthy controls (PERMANOVA, Table S19). **(H)** Density plots show the significant Gene

1097 Set Enrichments [142] of differentially abundant PICRUSt2 based KOs, between controls, CD  
1098 patients, and UC patients ( $P_{FDR} \leq 0.05$ ). Differential abundance was tested via *DESeq2* and the  
1099 enrichment score was derived from the  $-\log_{10}(\text{P-values}) \times \text{direction of fold change}$  (Table  
1100 S23). Repeatedly detected metabolic pathways are highlighted in bold.

1101

1102 **Figure 5: (A)** Spiec-Easi networks of baseline samples (N=1812). Bacterial nodes highlight  
1103 significant differentially abundant ASVs in the network [157]. Bacteria not showing any  
1104 differential abundance patterns between IBD patients and healthy controls are signified via  
1105 (●), bacteria overabundant in IBD (combined CD, UC, uIBD) via (■) and bacteria more  
1106 abundant in controls are signified via (▶). The barplot visualizes node centrality based on the  
1107 number of connections (node degree) at the baseline time point. Colored boxes highlight  
1108 corresponding ASV differential abundance patterns in KINDRED, external cohorts, and their  
1109 association to IBD onset and remission. Significance of centralities is derived from Z-tests  
1110 against a randomized networks ( $10^5$ 000). **(B)** Global network characteristics were derived  
1111 from the baseline and disease specific sub-networks, which can be informative for stability  
1112 and structure of the respective networks (Wilcoxon-test healthy vs. CD/UC). Network metrics  
1113 include centrality based assortativity [150], network diameter, radius, and size as well as  
1114 density/clustering [151,152], and natural connectivity [48]. **(C)** Network similarity of disease  
1115 and time point specific sub networks, as well as subnetworks of external cohorts (Malta,  
1116 Sweden) based on graphlet distance [153] and displayed via NMDS. Networks show a clear  
1117 compositional difference between healthy and diseased networks (Control vs. IBD (incl. IBD  
1118 networks):  $F_{1,18}=2.4144$ ,  $P=0.0412$ ,  $R^2=0.1183$ ,  $adj. R^2=0.0693$ ; PERMANOVA). **(D)**  
1119 Heritability estimates derived from the likelihood based method *lme4qtl* (Ziyatdinov *et al.*,  
1120 2018) using either only kinship information with or without additional environmental and  
1121 anthropometric covariates (▲-incl. environmental covariates, ▼-no covariates). Only the

1122 upper quartile of taxa are highlighted (based on  $h^2_{full}$  estimate including environmental  
1123 covariates). Additional information like differential abundance in IBD across cohorts are  
1124 depicted for each taxon (Figure S28A, Table S26).

1125

1126 **TABLES**

1127 **Table 1:** Baseline characteristics of the IBD patients ( $N_{\text{individuals}}=1868$ ) of the Kiel IBD Family Cohort. Values are mean  $\pm$  standard deviation or absolute  
 1128 counts. Abbreviations: CD, Crohn's disease; IBD, inflammatory bowel disease; UC, Ulcerative colitis, uIBD, undefined inflammatory bowel disease.

1129

Characteristics	Baseline				Follow-up 1				Follow-up 2			
	Control	CD	UC	uIBD	Control	CD	UC	uIBD	Control	CD	UC	uIBD
# Participants	791	551	438	32	362	174	105	7	295	146	92	6
Sex												
Male	316	175	166	13	140	57	38	1	103	52	34	0
Female	432	375	272	19	185	117	67	6	152	94	56	6
Age at baseline assessment, years	44.14 $\pm$ 1.94	43.99 $\pm$ 1.58	46.42 $\pm$ 1.62	41.63 $\pm$ 1.11	45.60 $\pm$ 1.03	43.95 $\pm$ 1.63	48.64 $\pm$ 1.37	49.71 $\pm$ 1.72	47.56 $\pm$ 1.06	47.03 $\pm$ 1.23	48.65 $\pm$ 1.82	55 $\pm$ 15.18
Age groups												
Children (7-11 years)	46	6	0	0	28	1	0	0	8	1	0	0
Adolescents (12-17 years)	47	24	8	2	33	11	0	0	11	3	0	0
Adults ( $\geq$ 18 years)	698	521	430	30	333	162	105	7	286	142	92	6
BMI (kg/m <sup>2</sup> ) average	24.98 $\pm$ 5.33	24.05 $\pm$ 4.74	24.60 $\pm$ 4.65	24.92 $\pm$ 4.88	25.18 $\pm$ 4.98	24.51 $\pm$ 5.28	24.65 $\pm$ 4.27	26.86 $\pm$ 3.57	26.15 $\pm$ 12.33	24.45 $\pm$ 4.71	24.99 $\pm$ 4.09	24.00 $\pm$ 5.30
BMI Children (7-11 yrs)	16.50 $\pm$ 2.47	14.78 $\pm$ 1.83	-	-	17.97 $\pm$ 2.68	14.18	-	-	16.89 $\pm$ 0.24	16.66	-	-
BMI Adolescents (12-17 yrs)	20.18 $\pm$ 3.	18.81 $\pm$ 2.5	17.11 $\pm$ 1.7	22.65 $\pm$ 3.1	20.66 $\pm$ 2.4	18.85 $\pm$ 2.6	-	-	22.37 $\pm$ 4.6	17.23 $\pm$ 1.1	-	-

BMI Adults (≥18 years)	18 25.81±4.92	4 24.36±4.60	2 24.74±4.58	3 25.07±4.98	0 25.70±4.81	4 24.96±5.15	24.65±4.27	26.86±3.57	2 26.36±12.48	5 24.66±4.61	24.99±4.09	24.00±5.30
Smoking Status												
never smoked	392	260	200	18	181	83	50	4	147	71	46	4
smoked less than 3 months	65	29	32		26	9	8	-	15	5	4	0
smoked in the past	201	185	181	12	84	62	39	2	70	55	30	1
smoking now	87	76	25	1	30	19	6	1	20	14	8	0
Age at IBD diagnosis (yrs)	-	24.42±12.22	29.94±13.66	23.1±18.68	-	22.51±10.51	26.74±10.14	71	-	23.63±11.17	28.06±13.77	33.33±33.72
Disease location for CD (multiple sites possible)												
not known	-	11	-	1	-	1	1	2	1	3	-	1
middle GI tract (proximal)	-	4	-	-	-	4	-	-	-	1	-	-
middle GI tract (distal)	-	9	1	-	-	6	-	-	-	2	-	-
upper GI tract	-	15	-	-	1	11	-	-	1	10	-	-
ileum	-	44	-	-	-	23	-	-	2	12	-	-
terminal ileum	2	145	1	2	-	67	1	-	8	51	1	1
ileum and colon	1	86	2	1	-	49	-	1	2	44	-	-
colon	2	75	16	6	-	44	6	3	2	32	-	3



perianal	30	1	1	-	15	-	-	1	8	-	-	
Disease location for UC (multiple sites possible)												
not known	-	-	6	1	-	1	1	6	2	-	2	1
proctitis	-	-	43	-	1	2	12	-	3	-	16	-
left colon	1	-	83	1	-	1	33	-	5	-	23	-
pancolitis	2	-	128	3	-	3	47	-	2	1	33	-
Medication												
Yes	479	536	417	30	205	167	99	6	163	130	80	6
No	312	15	21	2	157	7	6	1	132	16	12	0
antidiarrhetics	0	57	41	2	0	29	11	0	5	34	7	0
antibiotics	133	142	99	10	67	41	27	2	28	39	18	1
antiinflammatory	7	470	364	27	2	155	84	6	14	103	59	5
Type of IBD medication (last 12 months)												
Immunosuppressives	3	96	78	4	1	47	16	1	5	32	17	1
Glucocorticoids	1	39	29	0	0	29	15	3	4	33	12	0
5ASA	3	30	109	6	0	43	50	4	7	37	46	5

Mesalazin	0	117	112	11	0	49	57	4	-	-	-	
Sulfasalazin	0	24	16	0	0	12	4	0	-	-	-	
AZA	0	99	42	5	0	40	13	1	-	-	-	
Biologicals/‘small molecules’	1	77	45	1	0	54	16	0	4	46	12	0
IBD surgery												
Yes	2	364	78	4	0	116	83	1	0	98	19	0
No	789	187	360	28	362	58	22	6	295	48	73	6
Number of future onsets (IBD next time point)	4	-	-	-	1	-	-	-	2	-	-	-

1130

1131

1132 **Table 2:** Analysis of major physiological characteristics across time points using linear models. Results depicted are models after variable selection  
1133 minimizing AIC.

TP	Trait	Model Factors	DF	F	P	adj R2	Comparison	Mean		Slope	
								P	P(FDR)	P	P(FDR)
BL	ASCA IgA	BMI#	1,750	3.5792	0.0589	0.1835	CD - Contr.	$7.50 \times 10^{-33}$	$4.50 \times 10^{-32}$	-	-
	(log)	IBD	3,750	56.6106	$2.00 \times 10^{-16}$		CD - uIBD	$7.06 \times 10^{-6}$	$1.41 \times 10^{-5}$	-	-
							CD - UC	$1.54 \times 10^{-18}$	$4.62 \times 10^{-18}$	-	-

						Contr. - uIBD	0.6907	0.6907	-	-
						Contr. - UC	0.3482	0.5223	-	-
						uIBD - UC	0.4621	0.5545	-	-
ASCA IgG	Age#	1,775	5.5172	0.0191	0.1684	CD - Contr.	$3.26 \times 10^{-28}$	$1.95 \times 10^{-27}$	-	-
(log)	IBD	3,775	52.0840	$2.00 \times 10^{-16}$		CD - uIBD	$2.62 \times 10^{-6}$	$5.25 \times 10^{-6}$	-	-
						CD - UC	$5.15 \times 10^{-21}$	$1.55 \times 10^{-20}$	-	-
						Contr. - uIBD	0.3032	0.4548	-	-
						Contr. - UC	0.4175	0.4871	-	-
						uIBD - UC	0.4871	0.4871	-	-
GP2 IgA	Sex	1,744	0.0869	0.7682	0.0905	CD - Contr.	$6.98 \times 10^{-14}$	$4.19 \times 10^{-13}$	0.5056	0.7584
(log)	Age#	1,744	4.3479	0.0374		CD - uIBD	0.0808	0.1616	0.0099	0.0198
	BMI#	1,744	0.4423	0.5062		CD - UC	$9.09 \times 10^{-7}$	$2.73 \times 10^{-6}$	0.8492	0.8492
	IBD	3,744	20.5139	$8.85 \times 10^{-13}$		Contr. - uIBD	0.4227	0.5073	0.0040	0.0198
	Sex:Age#	1,744	10.1477	0.0015		Contr. - UC	0.2299	0.3449	0.7278	0.8492
	BMI#:IBD	3,744	2.8203	0.0381		uIBD - UC	0.7453	0.7453	0.0092	0.0198
GP2 IgG	Sex	1,770	1.3630	0.2434	0.0480	CD - Contr.	$2.84 \times 10^{-8}$	$1.71 \times 10^{-7}$	-	-
(log)	Age#	1,770	0.3528	0.5527		CD - uIBD	0.1108	0.2216	-	-
	IBD	3,770	11.3671	$2.67 \times 10^{-7}$		CD - UC	0.0001	0.0003	-	-

	Sex:Age#	1,770	9.3248	0.0023		Contr. - uIBD	0.8118	0.9654	-	-
						Contr. - UC	0.6145	0.9218	-	-
						uIBD - UC	0.9654	0.9654	-	-
CRP ( $X^{1/4}$ )	Age#	1,295	0.2114	0.6460	0.0323		-	-	-	-
(only CD & UC)	BMI#	1,295	9.7144	0.0020			-	-	-	-
	IBD	1,295	0.0050	0.9439			-	-	-	-
	Age#:IBD	1,295	4.0463	0.0452			-	-	-	-
HB ( $X^2$ )	Sex	1,414	123.8936	$2.36 \times 10^{-25}$	0.2408		-	-	-	-
(CD & UC)	BMI#	1,414	10.0727	0.0016			-	-	-	-
Calprotectin (rel.)	Age#	1,1755	6.5053	0.0108	0.0576	CD - Contr.	$7.58 \times 10^{-21}$	$4.55 \times 10^{-20}$	0.0091	0.0547
	IBD	3,1755	33.5156	$6.19 \times 10^{-22}$		CD - uIBD	0.4827	0.5792	0.9471	0.9471
	Age#:IBD	3,1755	2.5284	0.0558		CD - UC	0.0240	0.0431	0.4795	0.7690
						Contr. - uIBD	0.0287	0.0431	0.5127	0.7690
						Contr. - UC	$2.88 \times 10^{-10}$	$8.64 \times 10^{-10}$	0.1388	0.4164
						uIBD - UC	0.9287	0.9287	0.8561	0.9471
Bristol stool score	Age#	1,874	33.4743	$1.01 \times 10^{-8}$	0.1825	CD - Contr.	$1.50 \times 10^{-31}$	$8.97 \times 10^{-31}$	0.0013	0.0076
	IBD	3,874	53.2308	$2.20 \times 10^{-16}$		CD - uIBD	0.0682	0.0861	0.6533	0.9478
	Age#:IBD	3,874	3.4956	0.0152		CD - UC	0.0031	0.0063	0.0772	0.2315

							Contr. - uIBD	0.0718	0.0861	0.7898	0.9478
							Contr. - UC	1.94×10 <sup>-18</sup>	5.81×10 <sup>-18</sup>	0.2316	0.4632
							uIBD - UC	0.3532	0.3532	0.9902	0.9902
F1	CRP (X <sup>1/4</sup> , CD & UC)	Sex	1,175	2.2744	0.1333	0.0072		-	-	-	-
	HB (X <sup>2</sup> , CD & UC)	Sex	1,197	38.7355	2.86×10 <sup>-9</sup>	0.1601		-	-	-	-
	Calprotectin (rel.)	Age#	1,507	3.7867	0.0522	0.0921	CD - Contr.	4.54×10 <sup>-9</sup>	2.73×10 <sup>-8</sup>	0.0194	0.0583
		IBD	3,507	14.4469	4.81×10 <sup>-9</sup>		CD - uIBD	0.6170	0.7404	0.0860	0.1290
		Age#:IBD	3,507	4.0004	0.0078		CD - UC	0.5625	0.7404	0.5011	0.5011
							Contr. - uIBD	0.2880	0.5760	0.2935	0.3521
							Contr. - UC	2.91×10 <sup>-5</sup>	0.0001	0.0102	0.0583
							uIBD - UC	0.7708	0.7708	0.0508	0.1015
	Bristol stool score	Sex	1,314	4.2960	0.0390	0.0104		-	-	-	-
	QOL (relative)	Sex	1,593	5.1103	0.0242	0.1588	CD - Contr.	8.28×10 <sup>-18</sup>	4.97×10 <sup>-17</sup>	-	-
		Age#	1,593	23.4394	1.65×10 <sup>-6</sup>		CD - uIBD	0.1088	0.1632	-	-
		BMI#	1,593	5.4685	0.0197		CD - UC	0.0016	0.0031	-	-
		IBD	3,593	26.7214	3.20×10 <sup>-16</sup>		Contr. - uIBD	0.5548	0.5684	-	-
		Sex:BMI#	1,593	6.0564	0.0141		Contr. - UC	0.0001	0.0002	-	-
							uIBD - UC	0.5684	0.5684	-	-

F2	CRP ( $X^{1/4}$ , CD & UC)	Sex	1,106	0.3282	0.5679	0.0195	-	-	-	-	
		BMI#	1,106	0.2690	0.6051		-	-	-	-	
		Sex:BMI#	1,106	4.5680	0.0349		-	-	-	-	
	HB ( $X^2$ , CD & UC)	IBD	1,142	3.7767	0.0539	0.0191	-	-	-	-	
	Calprotectin (rel.)	IBD	2,144	6.4245	0.0021	0.0692	CD - Contr.	0.0009	0.0026	-	-
							CD - UC	0.2373	0.2373	-	-
							Contr. - UC	0.1120	0.1680	-	-
	Bristol stool score	Sex	1,250	1.0378	0.3093	0.0173	-	-	-	-	
		Age#	1,250	0.7100	0.4003		-	-	-	-	
		Sex:Age#	1,250	5.6996	0.0177		-	-	-	-	
	QOL (relative)	Age#	1,482	31.4038	$3.53 \times 10^{-8}$	0.1544	CD - Contr.	$2.99 \times 10^{-12}$	$1.80 \times 10^{-11}$	0.0756	0.0908
		BMI#	1,482	2.0787	0.1500		CD - uIBD	0.4478	0.4478	0.0283	0.0566
		IBD	3,482	17.7663	$5.94 \times 10^{-11}$		CD - UC	0.0169	0.0270	0.5291	0.5291
		Age#:IBD	3,482	3.5629	0.0142		Contr. - uIBD	0.0180	0.0270	0.0692	0.0908
						Contr. - UC	0.0005	0.0016	0.0271	0.0566	
						uIBD - UC	0.1527	0.1833	0.0181	0.0566	
FSS	Sex	1,479	8.0896	0.0046	0.0199	-	-	-	-		
	Age#	1,479	0.0910	0.7630		-	-	-	-		

	Sex:Age#	1,479	4.5930	0.0326	-	-	-	-
--	----------	-------	--------	--------	---	---	---	---

1134 # centered and scaled

1135 **Table 3:** Analysis of alpha diversity patterns across time points, correcting for age, sex, and  
 1136 BMI.

	Alpha diversity	Model	DF	F	P	adj. R2	Pairwise	
							Contrast	P (adjusted)
BL	Shannon (effective)	Age#	1,1727	4.1607	0.0415	0.1214	<b>CD-Contr.</b>	<b>4.95×10<sup>-12</sup></b>
		BMI#	1,1727	0.0108	0.9174		<b>CD-IBD</b>	<b>0.0483</b>
		Sex	1,1727	2.0196	0.1555		<b>CD-UC</b>	<b>5.16×10<sup>-6</sup></b>
		<b>IBD diagnosis</b>	3,1727	<b>79.7519</b>	<b>2.56×10<sup>-48</sup></b>		Contr.-IBD	0.1270
							<b>Contr.-UC</b>	<b>4.99×10<sup>-12</sup></b>
							IBD-UC	0.8411
	Chao	Age#	1,1727	4.1801	0.0411	0.1189	<b>CD-Contr.</b>	<b>4.95×10<sup>-12</sup></b>
		BMI#	1,1727	0.0440	0.8338		CD-IBD	0.0921
		Sex	1,1727	1.3698	0.2420		<b>CD-UC</b>	<b>4.18×10<sup>-7</sup></b>
		<b>IBD diagnosis</b>	3,1727	<b>78.0602</b>	<b>2.37×10<sup>-47</sup></b>		Contr.-IBD	0.0723
							<b>Contr.-UC</b>	<b>4.99×10<sup>-12</sup></b>
							IBD-UC	0.9774
unweighted NRI	Age#	1,1727	21.7176	0.0000	0.0146	CD-Contr.	0.1364	
	BMI#	1,1727	0.1274	0.7212		CD-IBD	0.7220	
	Sex	1,1727	0.7269	0.3940		CD-UC	0.9318	
	<b>IBD diagnosis</b>	3,1727	<b>3.0502</b>	<b>0.0276</b>		Contr.-IBD	0.9823	
						<b>Contr.-UC</b>	<b>0.0393</b>	
						IBD-UC	0.5945	
unweighted NTI	Age#	1,1727	18.6248	0.0000	0.0555	<b>CD-Contr.</b>	<b>6.52×10<sup>-12</sup></b>	
	BMI#	1,1727	8.1830	0.0043		CD-IBD	0.9981	
	Sex	1,1727	12.2254	0.0005		CD-UC	0.9177	
	<b>IBD diagnosis</b>	3,1727	<b>22.9073</b>	<b>1.53×10<sup>-14</sup></b>		Contr.-IBD	0.0588	



							<b>Contr.-UC</b>	<b>3.27×10<sup>-9</sup></b>
							IBD-UC	0.9783
F1	Shannon	Age#	1,599	3.7799	0.0523	0.1384	<b>CD-Contr.</b>	<b>4.54×10<sup>-10</sup></b>
	(effective)	BMI#	1,599	0.2665	0.6059		CD-IBD	0.7611
		Sex	1,599	3.8384	0.0506		<b>CD-UC</b>	<b>0.0026</b>
		<b>IBD diagnosis</b>	3,599	<b>31.7658</b>	<b>4.58×10<sup>-19</sup></b>		Contr.-IBD	0.4993
							<b>Contr.-UC</b>	<b>0.0002</b>
							IBD-UC	0.9988
	Chao	Age#	1,599	3.3731	0.0668	0.1091	<b>CD-Contr.</b>	<b>4.54×10<sup>-10</sup></b>
		BMI#	1,599	0.8586	0.3545		CD-IBD	0.8398
		Sex	1,599	0.3035	0.5819		<b>CD-UC</b>	<b>0.0081</b>
		<b>IBD diagnosis</b>	3,599	<b>25.1887</b>	<b>2.33×10<sup>-15</sup></b>		Contr.-IBD	0.5692
							<b>Contr.-UC</b>	<b>0.0013</b>
							IBD-UC	0.9975
	unweighted	Age#	1,599	2.3377	0.1268	0.0159	<b>CD-Contr.</b>	<b>0.0231</b>
	NRI	BMI#	1,599	4.1609	0.0418		CD-IBD	0.9981
		Sex	1,599	0.2207	0.6387		CD-UC	0.1840
		<b>IBD diagnosis</b>	3,599	<b>3.0230</b>	<b>0.0292</b>		Contr.-IBD	0.8141
							Contr.-UC	0.9979
							IBD-UC	0.8492
	unweighted	Age#	1,599	0.0932	0.7603	0.0657	<b>CD-Contr.</b>	<b>1.71×10<sup>-7</sup></b>
	NTI	BMI#	1,599	12.7048	0.0004		CD-IBD	0.7954
		Sex	1,599	0.3129	0.5761		<b>CD-UC</b>	<b>0.0060</b>
		<b>IBD diagnosis</b>	3,599	<b>11.8031</b>	<b>1.61×10<sup>-7</sup></b>		Contr.-IBD	0.0948
							Contr.-UC	0.6806
							IBD-UC	0.2089
F2	Shannon	Age#	1,483	2.1681	0.1415	0.1375	<b>CD-Contr.</b>	<b>6.25×10<sup>-11</sup></b>

(effective)	BMI#	1,483	0.0739	0.7859		CD-IBD	0.5878
	Sex	1,483	0.0128	0.9098		<b>CD-UC</b>	<b>0.0080</b>
	<b>IBD diagnosis</b>	3,483	<b>27.2317</b>	<b>2.75×10<sup>-16</sup></b>		Contr.-IBD	0.7581
						<b>Contr.-UC</b>	<b>0.0003</b>
	IBD-UC					IBD-UC	0.9958
Chao	Age#	1,483	5.4159	0.0204	0.1281	<b>CD-Contr.</b>	<b>6.25×10<sup>-11</sup></b>
	BMI#	1,483	2.8540	0.0918		CD-IBD	0.7797
	Sex	1,483	0.3818	0.5369		<b>CD-UC</b>	<b>0.0010</b>
	<b>IBD diagnosis</b>	3,483	<b>23.0590</b>	<b>5.77×10<sup>-14</sup></b>		Contr.-IBD	0.6626
						<b>Contr.-UC</b>	<b>0.0197</b>
	IBD-UC					IBD-UC	0.9929
unweighted	Age#	1,483	3.5979	0.0584	0.0179	CD-Contr.	0.1956
	NRI	BMI#	1,483	2.1588	0.1424	CD-IBD	0.5201
	Sex	1,483	1.4772	0.2248		CD-UC	0.9997
	<b>IBD diagnosis</b>	3,483	<b>2.5548</b>	<b>0.0547</b>		Contr.-IBD	0.2385
						Contr.-UC	0.3879
	IBD-UC					IBD-UC	0.5123
unweighted	Age#	1,483	3.0392	0.0819	0.0231	CD-Contr.	0.0585
	NTI	BMI#	1,483	3.0373	0.0820	CD-IBD	0.5655
	Sex	1,483	0.0126	0.9105		CD-UC	0.9990
	<b>IBD diagnosis</b>	3,483	<b>3.8270</b>	<b>0.0099</b>		Contr.-IBD	0.2120
						Contr.-UC	0.1015
	IBD-UC					IBD-UC	0.6040

1137 # scaled and centered

1138

1139

1140

## 1141 REFERENCES

1. Burisch J, Jess T, Martinato M, Lakatos PL, on behalf of ECCO -EpiCom. The burden of inflammatory bowel disease in Europe. *J Crohns Colitis* [Internet]. 2013 [cited 2023 Mar 10];7:322–37. Available from: <https://doi.org/10.1016/j.crohns.2013.01.010>
2. Frolkis A, Dieleman LA, Barkema HW, Panaccione R, Ghosh S, Fedorak RN, et al. Environment and the inflammatory bowel diseases. *Can J Gastroenterol J Can Gastroenterol*. 2013;27:e18-24.
3. Nunes T, Fiorino G, Danese S, Sans M. Familial aggregation in inflammatory bowel disease: Is it genes or environment? *World J Gastroenterol WJG* [Internet]. 2011 [cited 2023 Mar 10];17:2715–22. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3123468/>
4. Zhao M, Burisch J. Impact of Genes and the Environment on the Pathogenesis and Disease Course of Inflammatory Bowel Disease. *Dig Dis Sci*. 2019;64:1759–69.
5. Molodecky NA, Soon IS, Rabi DM, Ghali WA, Ferris M, Chernoff G, et al. Increasing incidence and prevalence of the inflammatory bowel diseases with time, based on systematic review. *Gastroenterology* [Internet]. 2012;142:46-54.e42; quiz e30. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/22001864>
6. Alatab S, Sepanlou SG, Ikuta K, Vahedi H, Bisignano C, Safiri S, et al. The global, regional, and national burden of inflammatory bowel disease in 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet Gastroenterol Hepatol* [Internet]. 2020 [cited 2023 Mar 10];5:17–30. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S2468125319303334>

7. Ng SC, Shi HY, Hamidi N, Underwood FE, Tang W, Benchimol EI, et al. Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. *The Lancet* [Internet]. 2017 [cited 2023 Mar 10];390:2769–78. Available from: [https://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(17\)324480/fulltext](https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(17)324480/fulltext)
8. Kaplan GG, Windsor JW. The four epidemiological stages in the global evolution of inflammatory bowel disease. *Nat Rev Gastroenterol Hepatol*. 2021;18:56–66.
9. Schreiber S, Rosenstiel P, Franke A. [Genetic etiology of chronic inflammatory bowel disease]. *Internist*. 2014;55:156–64.
10. Rosenstiel P, Sina C, Franke A, Schreiber S. Towards a molecular risk map—Recent advances on the etiology of inflammatory bowel disease. *Semin Immunol* [Internet]. 2009 [cited 2023 Mar 10];21:334–45. Available from: <https://www.sciencedirect.com/science/article/pii/S1044532309000955>
11. Moller FT, Andersen V, Wohlfahrt J, Jess T. Familial risk of inflammatory bowel disease: a population-based cohort study 1977-2011. *Am J Gastroenterol*. 2015;110:564–71.
12. Capone K, Rosenberg HJ, Wroblewski K, Gokhale R, Kirschner BS. Change in Prevalence of Family History During Long-term Follow-up of Patients With Pediatric-onset Inflammatory Bowel Disease. *J Pediatr Gastroenterol Nutr*. 2019;68:829–34.
13. Orholm M, Munkholm P, Langholz E, Nielsen OH, Sørensen TIA, Binder V. Familial Occurrence of Inflammatory Bowel Disease. *N Engl J Med*. 1991;324:84–8.
14. Santos MPC, Gomes C, Torres J. Familial and ethnic risk in inflammatory bowel disease. *Ann Gastroenterol*. 2018;31:14–23.

15. Schreiber S, Rosenstiel P, Albrecht M, Hampe J, Krawczak M. Genetics of Crohn disease, an archetypal inflammatory barrier disease. *Nat Rev Genet* [Internet]. 2005;6:376–88.

Available from: <http://dx.doi.org/10.1038/nrg1607>

16. Liu Z, Liu R, Gao H, Jung S, Gao X, Sun R, et al. Genetic architecture of the inflammatory bowel diseases across East Asian and European ancestries. *Nat Genet* [Internet]. 2023 [cited 2024 May 13];55:796–806. Available from:

<https://www.nature.com/articles/s41588-023-01384-0>

17. Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, Hui KY, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature*. 2012;491:119–24.

18. Anderson CA, Boucher G, Lees CW, Franke A, D’Amato M, Taylor KD, et al. Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. *Nat Genet*. 2011;43:246–52.

19. Cleynen I, Boucher G, Jostins L, Schumm LP, Zeissig S, Ahmad T, et al. Inherited determinants of Crohn’s disease and ulcerative colitis phenotypes: a genetic association study. *The Lancet* [Internet]. 2016 [cited 2023 Mar 10];387:156–67. Available from:

[https://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(15\)00465-1/fulltext?rss%3Dyes=](https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(15)00465-1/fulltext?rss%3Dyes=)

20. Liu JZ, van Sommeren S, Huang H, Ng SC, Alberts R, Takahashi A, et al. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat Genet* [Internet]. 2015 [cited 2023 Mar 10];47:979–86.

Available from: <https://www.nature.com/articles/ng.3359>

21. Ballester MP, Martí D, Tosca J, Bosca-Watts MM, Sanahuja A, Navarro P, et al. Disease severity and treatment requirements in familial inflammatory bowel disease. *Int J Colorectal Dis.* 2017;32:1197–205.
22. Borren NZ, Conway G, Garber JJ, Khalili H, Budree S, Mallick H, et al. Differences in Clinical Course, Genetics, and the Microbiome Between Familial and Sporadic Inflammatory Bowel Diseases. *J Crohns Colitis.* 2018;12:525–31.
23. Kevans D, Silverberg MS, Borowski K, Griffiths A, Xu W, Onay V, et al. IBD Genetic Risk Profile in Healthy First-Degree Relatives of Crohn’s Disease Patients. *J Crohns Colitis.* 2016;10:209–15.
24. Torkamani A, Wineinger NE, Topol EJ. The personal and clinical utility of polygenic risk scores. *Nat Rev Genet [Internet].* 2018 [cited 2024 Apr 11];19:581–90. Available from: <https://www.nature.com/articles/s41576-018-0018-x>
25. Lloyd-Price J, Arze C, Ananthakrishnan AN, Schirmer M, Avila-Pacheco J, Poon TW, et al. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature [Internet].* 2019;569:655–62. Available from: <https://doi.org/10.1038/s41586-019-1237-9>
26. Proctor LM, Creasy HH, Fettweis JM, Lloyd-Price J, Mahurkar A, Zhou W, et al. The Integrative Human Microbiome Project. *Nature [Internet].* 2019;569:641–8. Available from: <https://doi.org/10.1038/s41586-019-1238-8>
27. Lewis JD, Chen EZ, Baldassano RN, Otley AR, Griffiths AM, Lee D, et al. Inflammation, Antibiotics, and Diet as Environmental Stressors of the Gut Microbiome in Pediatric Crohn’s Disease. *Cell Host Microbe [Internet].* 2015;18:489–500. Available from: <file://www.sciencedirect.com/science/article/pii/S1931312815003777>

28. Torres J, Burisch J, Riddle M, Dubinsky M, Colombel J-F. Preclinical disease and preventive strategies in IBD: perspectives, challenges and opportunities. *Gut* [Internet]. 2016 [cited 2024 Apr 19];65:1061–9. Available from: <https://gut.bmj.com/content/65/7/1061>
29. Argmann C, Hou R, Ungaro RC, Irizar H, Al-Taie Z, Huang R, et al. Biopsy and blood-based molecular biomarker of inflammation in IBD. *Gut* [Internet]. 2022;0:1–17. Available from: <https://gut.bmj.com/lookup/doi/10.1136/gutjnl-2021-326451>
30. Torres J, Petralia F, Sato T, Wang P, Telesco SE, Choung RS, et al. Serum Biomarkers Identify Patients Who Will Develop Inflammatory Bowel Diseases Up to 5 Years Before Diagnosis. *Gastroenterology* [Internet]. 2020;159:96–104. Available from: <https://doi.org/10.1053/j.gastro.2020.03.007>
31. Leibovitch H, Lee S-H, Garay JAR, Espin-Garcia O, Xue M, Neustaeter A, et al. Immune response and barrier dysfunction-related proteomic signatures in preclinical phase of Crohn's disease highlight earliest events of pathogenesis. *Gut* [Internet]. 2023 [cited 2023 Aug 31];72:1462–71. Available from: <https://gut.bmj.com/content/72/8/1462>
32. Galipeau HJ, Caminero A, Turpin W, Bermudez-Brito M, Santiago A, Libertucci J, et al. Novel Fecal Biomarkers That Precede Clinical Diagnosis of Ulcerative Colitis. *Gastroenterology*. 2021;160:1532–45.
33. Roggenbuck D, Hausdorf G, Martinez-Gamboa L, Reinhold D, Büttner T, Jungblut PR, et al. Identification of GP2, the major zymogen granule membrane glycoprotein, as the autoantigen of pancreatic antibodies in Crohn's disease. *Gut* [Internet]. 2009 [cited 2024 Jun 6];58:1620–8. Available from: <https://gut.bmj.com/content/58/12/1620>

34. Roggenbuck D, Reinhold D, Schierack P, Bogdanos DP, Conrad K, Laass MW. Crohn's disease specific pancreatic antibodies: clinical and pathophysiological challenges. *Clin Chem Lab Med*. 2014;52:483–94.
35. Jendrek ST, Gotthardt D, Nitzsche T, Widmann L, Korf T, Michaels MA, et al. Anti-GP2 IgA autoantibodies are associated with poor survival and cholangiocarcinoma in primary sclerosing cholangitis. *Gut*. 2017;66:137–44.
36. Privé F, Arbel J, Vilhjálmsson BJ. LDpred2: better, faster, stronger. *Bioinforma Oxf Engl*. 2021;36:5424–31.
37. Kuehbacher T, Rehman A, Lepage P, Hellmig S, Fölsch UR, Schreiber S, et al. Intestinal TM7 bacterial phylogenies in active inflammatory bowel disease. *J Med Microbiol*. 2008;57:1569–76.
38. Gevers D, Kugathasan S, Denson LA, Vázquez-Baeza Y, Van Treuren W, Ren B, et al. The treatment-naive microbiome in new-onset Crohn's disease. *Cell Host Microbe*. 2014;15:382–92.
39. Gupta VK, Kim M, Bakshi U, Cunningham KY, Davis JM, Lazaridis KN, et al. A predictive index for health status using species-level gut microbiome profiling. *Nat Commun* [Internet]. 2020;11. Available from: <http://dx.doi.org/10.1038/s41467-020-18476-8>
40. Maier L, Pruteanu M, Kuhn M, Zeller G, Telzerow A, Anderson EE, et al. Extensive impact of non-antibiotic drugs on human gut bacteria. *Nature* [Internet]. 2018;555:623–8. Available from: <http://dx.doi.org/10.1038/nature25979>
41. Bennett M, Gilroy DW. Lipid Mediators in Inflammation. *Microbiol Spectr* [Internet]. 2016 [cited 2023 May 26];4:4.6.06. Available from: <https://journals.asm.org/doi/10.1128/microbiolspec.MCHD-0035-2016>



42. Benus RFJ, Van Der Werf TS, Welling GW, Judd PA, Taylor MA, Harmsen HJM, et al. Association between *Faecalibacterium prausnitzii* and dietary fibre in colonic fermentation in healthy human subjects. *Br J Nutr* [Internet]. 2010;104:693–700. Available from: <http://dx.doi.org/10.1017/S0007114510001030>
43. Vieira-Silva S, Sabino J, Valles-Colomer M, Falony G, Kathagen G, Caenepeel C, et al. Quantitative microbiome profiling disentangles inflammation- and bile duct obstruction-associated microbiota alterations across PSC/IBD diagnoses. *Nat Microbiol* [Internet]. 2019 [cited 2023 May 4];4:1826–31. Available from: <https://www.nature.com/articles/s41564-019-0483-9>
44. Zhu L, Baker SS, Gill C, Liu W, Alkhoury R, Baker RD, et al. Characterization of gut microbiomes in nonalcoholic steatohepatitis (NASH) patients: a connection between endogenous alcohol and NASH. *Hepatol Baltim Md*. 2013;57:601–9.
45. Costea PI, Hildebrand F, Manimozhayan A, Bäckhed F, Blaser MJ, Bushman FD, et al. Enterotypes in the landscape of gut microbial community composition. *Nat Microbiol* [Internet]. 2017;3:8–16. Available from: <https://doi.org/10.1038/s41564-017-0072-8>
46. Frioux C, Ansoorge R, Özkurt E, Ghassemi Nedjad C, Fritscher J, Quince C, et al. Enterosignatures define common bacterial guilds in the human gut microbiome. *Cell Host Microbe* [Internet]. 2023 [cited 2023 Jun 30]; Available from: <https://www.sciencedirect.com/science/article/pii/S1931312823002172>
47. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102:15545–50.

48. Jun W, Barahona M, Yue-Jin T, Hong-Zhong D. Natural Connectivity of Complex Networks. *Chin Phys Lett* [Internet]. 2010 [cited 2023 Aug 27];27:078902. Available from: <https://dx.doi.org/10.1088/0256-307X/27/7/078902>
49. Oehlers M, Fabian B. Graph Metrics for Network Robustness—A Survey. *Mathematics* [Internet]. 2021 [cited 2023 Jul 19];9:895. Available from: <https://www.mdpi.com/2227-7390/9/8/895>
50. Robertson LW, Chandrasekaran A, Reuning RH, Hui J, Rawal BD. Reduction of digoxin to 20R-dihydrodigoxin by cultures of *Eubacterium lentum*. *Appl Environ Microbiol* [Internet]. 1986 [cited 2023 May 13];51:1300–3. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC239061/>
51. Salaritabar A, Darvishi B, Hadjiakhoondi F, Manayi A, Sureda A, Nabavi SF, et al. Therapeutic potential of flavonoids in inflammatory bowel disease: A comprehensive review. *World J Gastroenterol* [Internet]. 2017 [cited 2023 Oct 26];23:5097–114. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5537178/>
52. Gupta A, Dhakan DB, Maji A, Saxena R, P.K. VP, Mahajan S, et al. Association of *Flavonifractor plautii*, a Flavonoid-Degrading Bacterium, with the Gut Microbiome of Colorectal Cancer Patients in India. *mSystems*. 2019;4:e00438-19.
53. Moll JM, Myers PN, Zhang C, Eriksen C, Wolf J, Appelberg KS, et al. Gut Microbiota Perturbation in IgA Deficiency Is Influenced by IgA-Autoantibody Status. *Gastroenterology* [Internet]. 2021 [cited 2024 Feb 20];160:2423-2434.e5. Available from: [https://www.gastrojournal.org/article/S0016-5085\(21\)00458-3/fulltext](https://www.gastrojournal.org/article/S0016-5085(21)00458-3/fulltext)
54. Alexander M, Ang QY, Nayak RR, Bustion AE, Sandy M, Zhang B, et al. Human gut bacterial metabolism drives Th17 activation and colitis. *Cell Host Microbe*. 2022;30:17-30.e9.

55. Banerjee S, Schlaeppi K, van der Heijden MGA. Keystone taxa as drivers of microbiome structure and functioning. *Nat Rev Microbiol* [Internet]. 2018 [cited 2024 Apr 17];16:567–76. Available from: <https://www.nature.com/articles/s41579-018-0024-1>
56. Röttgers L, Faust K. Can we predict keystones? *Nat Rev Microbiol*. 2019;17:193.
57. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, et al. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* [Internet]. 2010;464:59–65. Available from: <http://dx.doi.org/10.1038/nature08821>
58. Gordon H, Trier Moller F, Andersen V, Harbord M. Heritability in Inflammatory Bowel Disease: From the First Twin Study to Genome-Wide Association Studies. *Inflamm Bowel Dis* [Internet]. 2015 [cited 2023 May 3];21:1428–34. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4450891/>
59. Turpin W, Espin-Garcia O, Xu W, Silverberg MS, Kevans D, Smith MI, et al. Association of host genome with intestinal microbial composition in a large healthy cohort. *Nat Genet* [Internet]. 2016;48:1413–7. Available from: <http://dx.doi.org/10.1038/ng.3693>
60. Rausch P, Rehman A, Künzel S, Häsler R, Ott SJ, Schreiber S, et al. Colonic mucosa-associated microbiota is influenced by an interaction of crohn disease and FUT2 (Secretor) genotype. *Proc Natl Acad Sci U S A* [Internet]. 2011;108:19030–5. Available from: <http://www.pnas.org/content/early/2011/11/08/1106408108.abstract>
61. Zaneveld JR, McMinds R, Thurber RV. Stress and stability: Applying the Anna Karenina principle to animal microbiomes. *Nat Microbiol* [Internet]. 2017;2:17121. Available from: <https://doi.org/10.1038/nmicrobiol.2017.121>

62. Halfvarson J, Brislawn CJ, Lamendella R, Vázquez-Baeza Y, Walters WA, Bramer LM, et al. Dynamics of the human gut microbiome in inflammatory bowel disease. *Nat Microbiol* [Internet]. 2017 [cited 2023 Jan 18];2. Available from: </pmc/articles/PMC5319707/>
63. Pisani A, Rausch P, Bang C, Ellul S, Tabone T, Marantidis Cordina C, et al. Dysbiosis in the Gut Microbiota in Patients with Inflammatory Bowel Disease during Remission. Auchtung JM, editor. *Microbiol Spectr* [Internet]. 2022;10:e0061622. Available from: <https://journals.asm.org/doi/10.1128/spectrum.00616-22>
64. Rausch P, Ellul S, Pisani A, Bang C, Tabone T, Marantidis Cordina C, et al. Microbial Dynamics in Newly Diagnosed and Treatment Naïve IBD Patients in the Mediterranean. *Inflamm Bowel Dis* [Internet]. 2023 [cited 2023 Feb 7];1–15. Available from: <https://academic.oup.com/ibdjournal/advance-article/doi/10.1093/ibd/izad004/7025776>
65. Halfvarson J, Standaert-Vitse A, Järnerot G, Sendid B, Jouault T, Bodin L, et al. Anti-*Saccharomyces cerevisiae* antibodies in twins with inflammatory bowel disease. *Gut* [Internet]. 2005 [cited 2022 Oct 6];54:1237–43. Available from: [www.gutjnl.com](http://www.gutjnl.com)
66. Schirmer M, Denson L, Vlamakis H, Franzosa EA, Thomas S, Gotman NM, et al. Compositional and Temporal Changes in the Gut Microbiome of Pediatric Ulcerative Colitis Patients Are Linked to Disease Course. *Cell Host Microbe* [Internet]. 2018;24:600-610.e4. Available from: <https://www.sciencedirect.com/science/article/pii/S1931312818304931>
67. Falony G, Joossens M, Vieira-Silva S, Wang J, Darzi Y, Faust K, et al. Population-level analysis of gut microbiome variation. *Science*. 2016;352:560–4.
68. Vandeputte D, Falony G, Vieira-Silva S, Tito RY, Joossens M, Raes J. Stool consistency is strongly associated with gut microbiota richness and composition, enterotypes and bacterial

growth rates. *Gut* [Internet]. 2016 [cited 2024 Jan 31];65:57–62. Available from:

<https://gut.bmj.com/content/65/1/57>

69. Boekhorst J, Venlet N, Procházková N, Hansen ML, Lieberoth CB, Bahl MI, et al. Stool energy density is positively correlated to intestinal transit time and related to microbial enterotypes. *Microbiome* [Internet]. 2022 [cited 2024 Apr 15];10:1–10. Available from:

<https://microbiomejournal.biomedcentral.com/articles/10.1186/s40168-022-01418-5>

70. Procházková N, Falony G, Dragsted LO, Licht TR, Raes J, Roager HM. Advancing human gut microbiota research by considering gut transit time. *Gut* [Internet]. 2023 [cited 2024 Apr 15];72:180–91. Available from: <https://gut.bmj.com/content/72/1/180>

71. Turpin W, Bedrani L, Espin-Garcia O, Xu W, Silverberg MS, Smith MI, et al. Associations of NOD2 polymorphisms with Erysipelotrichaceae in stool of in healthy first degree relatives of Crohn’s disease subjects. *BMC Med Genet* [Internet]. 2020 [cited 2023 May 14];21:1–8. Available from:

<https://bmcmmedgenet.biomedcentral.com/articles/10.1186/s12881-020-01115-w>

72. Ellinghaus D, Bethune J, Petersen B-S, Franke A. The genetics of Crohn’s disease and ulcerative colitis – status quo and beyond. *Scand J Gastroenterol* [Internet]. 2015 [cited 2024 Apr 19];50:13–23. Available from: <https://doi.org/10.3109/00365521.2014.990507>

73. Schirmer M, Stražar M, Avila-Pacheco J, Rojas-Tapias DF, Brown EM, Temple E, et al. Linking microbial genes to plasma and stool metabolites uncovers host-microbial interactions underlying ulcerative colitis disease course. *Cell Host Microbe* [Internet]. 2024 [cited 2024 Feb 19];32:209-226.e7. Available from:

<https://www.sciencedirect.com/science/article/pii/S1931312823005085>

74. Wegiel B, Otterbein L. Go Green: The Anti-Inflammatory Effects of Biliverdin Reductase. *Front Pharmacol* [Internet]. 2012 [cited 2024 Apr 11];3. Available from: <https://www.frontiersin.org/journals/pharmacology/articles/10.3389/fphar.2012.00047/full>
75. Schirmer M, Smeekens SP, Vlamakis H, Jaeger M, Oosting M, Franzosa EA, et al. Linking the Human Gut Microbiome to Inflammatory Cytokine Production Capacity. *Cell* [Internet]. 2016 [cited 2024 Feb 27];167:1125-1136.e8. Available from: [https://www.cell.com/cell/abstract/S0092-8674\(16\)31403-9](https://www.cell.com/cell/abstract/S0092-8674(16)31403-9)
76. Rühlemann MC, Hermes BM, Bang C, Doms S, Moitinho-Silva L, Thingholm LB, et al. Genome-wide association study in 8,956 German individuals identifies influence of ABO histo-blood groups on gut microbiome. *Nat Genet* [Internet]. 2021;53:147–55. Available from: <https://doi.org/10.1038/s41588-020-00747-1>
77. Mazmanian SK, Round JL, Kasper DL. A microbial symbiosis factor prevents intestinal inflammatory disease. *Nature* [Internet]. 2008 [cited 2024 Mar 14];453:620–5. Available from: <https://www.nature.com/articles/nature07008>
78. Chu H, Khosravi A, Kusumawardhani IP, Kwon AHK, Vasconcelos AC, Cunha LD, et al. Gene-microbiota interactions contribute to the pathogenesis of inflammatory bowel disease. *Science*. 2016;352:1116–20.
79. Goodrich JK, Davenport ER, Beaumont M, Jackson MA, Knight R, Ober C, et al. Genetic Determinants of the Gut Microbiome in UK Twins. *Cell Host Microbe* [Internet]. 2016;19:731–43. Available from: <http://dx.doi.org/10.1016/j.chom.2016.04.017>
80. Valles-Colomer M, Blanco-Míguez A, Manghi P, Asnicar F, Dubois L, Golzato D, et al. The person-to-person transmission landscape of the gut and oral microbiomes. *Nature* [Internet]. 2023;614:125–35. Available from: <https://doi.org/10.1038/s41586-022-05620-1>

81. Goodrich JK, Waters JL, Poole AC, Sutter JL, Koren O, Blekhman R, et al. Human genetics shape the gut microbiome. *Cell* [Internet]. 2014;159:789–99. Available from: <http://www.sciencedirect.com/science/article/pii/S0092867414012410>
82. Schulfer AF, Battaglia T, Alvarez Y, Bijnens L, Ruiz VE, Ho M, et al. Intergenerational transfer of antibiotic-perturbed microbiota enhances colitis in susceptible mice. *Nat Microbiol* [Internet]. 2018;3:234–42. Available from: <http://dx.doi.org/10.1038/s41564-017-0075-5>
83. Parker A, Romano S, Ansoorge R, Aboelnour A, Le Gall G, Savva GM, et al. Fecal microbiota transfer between young and aged mice reverses hallmarks of the aging gut, eye, and brain. *Microbiome* [Internet]. 2022 [cited 2023 Sep 21];10:1–25. Available from: <https://microbiomejournal.biomedcentral.com/articles/10.1186/s40168-022-01243-w>
84. Wang J, Kalyan S, Steck N, Turner LM, Harr B, Künzel S, et al. Analysis of intestinal microbiota in hybrid house mice reveals evolutionary divergence in a vertebrate hologenome. *Nat Commun*. 2015;6:6440.
85. Koropatkin NM, Cameron EA, Martens EC. How glycan metabolism shapes the human gut microbiota. *Nat Rev Microbiol* [Internet]. 2012;10:323–35. Available from: <http://dx.doi.org/10.1038/nrmicro2746>
86. Desai MS, Seekatz AM, Koropatkin NM, Kamada N, Hickey CA, Wolter M, et al. A Dietary Fiber-Deprived Gut Microbiota Degrades the Colonic Mucus Barrier and Enhances Pathogen Susceptibility. *Cell* [Internet]. 2016 [cited 2024 Apr 8];167:1339-1353.e21. Available from: <https://www.sciencedirect.com/science/article/pii/S0092867416314647>
87. Gacesa R, Kurilshikov A, Vich Vila A, Sinha T, Klaassen MAY, Bolte LA, et al. Environmental factors shaping the gut microbiome in a Dutch population. *Nature* [Internet]. 2022;604:732–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/35418674>

88. Rajilić-Stojanović M, Shanahan F, Guarner F, De Vos WM. Phylogenetic analysis of dysbiosis in ulcerative colitis during remission. *Inflamm Bowel Dis*. 2013;19:481–8.
89. Berry D, Schwab C, Milinovich G, Reichert J, Ben Mahfoudh K, Decker T, et al. Phylotype-level 16S rRNA analysis reveals new bacterial indicators of health state in acute murine colitis. *ISME J [Internet]*. 2012;6:2091–106. Available from: <http://dx.doi.org/10.1038/ismej.2012.39>
90. Waters JL, Ley RE. The human gut bacteria Christensenellaceae are widespread, heritable, and associated with health. *BMC Biol [Internet]*. 2019 [cited 2024 Apr 16];17:83. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6819567/>
91. Utter DR, He X, Cavanaugh CM, McLean JS, Bor B. The saccharibacterium TM7x elicits differential responses across its host range. *ISME J [Internet]*. 2020 [cited 2024 Apr 17];14:3054–67. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7784981/>
92. McLean JS, Bor B, Kerns KA, Liu Q, To TT, Solden L, et al. Acquisition and Adaptation of Ultra-small Parasitic Reduced Genome Bacteria to Mammalian Hosts. *Cell Rep [Internet]*. 2020;32:107939. Available from: <https://doi.org/10.1016/j.celrep.2020.107939>
93. Hong M, Li Z, Liu H, Zheng S, Zhang F, Zhu J, et al. *Fusobacterium nucleatum* aggravates rheumatoid arthritis through FadA-containing outer membrane vesicles. *Cell Host Microbe [Internet]*. 2023 [cited 2023 May 12];31:798-810.e7. Available from: [https://www.cell.com/cell-host-microbe/abstract/S1931-3128\(23\)00123-3](https://www.cell.com/cell-host-microbe/abstract/S1931-3128(23)00123-3)
94. Engevik MA, Danhof HA, Ruan W, Engevik AC, Chang-Graham AL, Engevik KA, et al. *Fusobacterium nucleatum* Secretes Outer Membrane Vesicles and Promotes Intestinal Inflammation. *mBio [Internet]*. 2021 [cited 2023 May 12];12:e02706-20. Available from: <https://journals.asm.org/doi/10.1128/mBio.02706-20>



95. Bullman S, Pedomallu CS, Sicinska E, Clancy TE, Zhang X, Cai D, et al. Analysis of *Fusobacterium* persistence and antibiotic response in colorectal cancer. *Science* [Internet]. 2017 [cited 2023 Dec 10];358:1443–8. Available from: <https://www.science.org/doi/10.1126/science.aal5240>
96. Rühlemann MC, Heinsen FA, Zenouzi R, Lieb W, Franke A, Schramm C. Faecal microbiota profiles as diagnostic biomarkers in primary sclerosing cholangitis. *Gut*. 2017;66:753–4.
97. Atarashi K, Suda W, Luo C, Kawaguchi T, Motoo I, Narushima S, et al. Ectopic colonization of oral bacteria in the intestine drives TH1 cell induction and inflammation. *Science*. 2017;358:359–65.
98. Madsen GR, Bertl K, Pandis N, Stavropoulos A, Burisch J. The Impact of Periodontitis on Inflammatory Bowel Disease Activity. *Inflamm Bowel Dis*. 2023;29:396–404.
99. Conte MP, Longhi C, Marazzato M, Conte AL, Aleandri M, Lepanto MS, et al. Adherent-invasive *Escherichia coli* (AIEC) in pediatric Crohn’s disease patients: Phenotypic and genetic pathogenic features. *BMC Res Notes* [Internet]. 2014;7:748. Available from: <https://doi.org/10.1186/1756-0500-7-748>
100. Stecher B, Chaffron S, Käppeli R, Hapfelmeier S, Friedrich S, Weber TC, et al. Like will to like: Abundances of closely related species can predict susceptibility to intestinal colonization by pathogenic and commensal bacteria. *PLoS Pathog* [Internet]. 2010;6:e1000711. Available from: <http://dx.doi.org/10.1371/journal.ppat.1000711>
101. Winter SE, Thiennimitr P, Winter MG, Butler BP, Huseby DL, Crawford RW, et al. Gut inflammation provides a respiratory electron acceptor for *Salmonella*. *Nature* [Internet]. 2010 [cited 2023 Dec 12];467:426–9. Available from: <https://www.nature.com/articles/nature09415>

102. Lupp C, Robertson ML, Wickham ME, Sekirov I, Champion OL, Gaynor EC, et al. Host-Mediated Inflammation Disrupts the Intestinal Microbiota and Promotes the Overgrowth of Enterobacteriaceae. *Cell Host Microbe* [Internet]. 2007;2:119–29. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S1931312807001576>
103. Vatanen T, Kostic AD, D’Hennezel E, Siljander H, Franzosa EA, Yassour M, et al. Variation in Microbiome LPS Immunogenicity Contributes to Autoimmunity in Humans. *Cell* [Internet]. 2016;165:842–53. Available from: <http://www.sciencedirect.com/science/article/pii/S0092867416303981>
104. Schultz M. Clinical use of E. coli Nissle 1917 in inflammatory bowel disease. *Inflamm Bowel Dis* [Internet]. 2008 [cited 2024 Apr 19];14:1012–8. Available from: <https://dx.doi.org/10.1002/ibd.20377>
105. Ma S, Shungin D, Mallick H, Schirmer M, Nguyen LH, Kolde R, et al. Population structure discovery in meta-analyzed microbial communities and inflammatory bowel disease using MMUPHin. *Genome Biol* [Internet]. 2022 [cited 2024 Jun 11];23:1–31. Available from: <https://genomebiology.biomedcentral.com/articles/10.1186/s13059-022-02753-4>
106. Zelante T, Iannitti RG, Cunha C, DeLuca A, Giovannini G, Pieraccini G, et al. Tryptophan catabolites from microbiota engage aryl hydrocarbon receptor and balance mucosal reactivity via interleukin-22. *Immunity* [Internet]. 2013;39:372–85. Available from: <http://dx.doi.org/10.1016/j.immuni.2013.08.003>
107. Nikolaus S, Schulte B, Al-Massad N, Thieme F, Schulte DM, Bethge J, et al. Increased Tryptophan Metabolism Is Associated With Activity of Inflammatory Bowel Diseases. *Gastroenterology*. 2017;153:1504-1516.e2.

108. Clasen SJ, Bell MEW, Borbón A, Lee D-H, Henseler ZM, de la Cuesta-Zuluaga J, et al. Silent recognition of flagellins from human gut commensal bacteria by Toll-like receptor 5. *Sci Immunol* [Internet]. 2023 [cited 2023 Sep 22];8:eabq7001. Available from: <https://www.science.org/doi/10.1126/sciimmunol.abq7001>
109. Alexander KL, Zhao Q, Reif M, Rosenberg AF, Mannon PJ, Duck LW, et al. Human Microbiota Flagellins Drive Adaptive Immune Responses in Crohn's Disease. *Gastroenterology* [Internet]. 2021 [cited 2024 Apr 18];161:522-535.e6. Available from: <https://www.sciencedirect.com/science/article/pii/S0016508521006338>
110. Lee S-H, Turpin W, Espin-Garcia O, Raygoza Garay JA, Smith MI, Leibovitz H, et al. Anti-Microbial Antibody Response is Associated With Future Onset of Crohn's Disease Independent of Biomarkers of Altered Gut Barrier Function, Subclinical Inflammation, and Genetic Risk. *Gastroenterology* [Internet]. 2021 [cited 2024 Apr 18];161:1540–51. Available from: <https://www.sciencedirect.com/science/article/pii/S0016508521032431>
111. Turpin W, Lee S-H, Raygoza Garay JA, Madsen KL, Meddings JB, Bedrani L, et al. Increased Intestinal Permeability Is Associated With Later Development of Crohn's Disease. *Gastroenterology* [Internet]. 2020 [cited 2023 Jun 14];159:2092-2100.e5. Available from: <https://www.sciencedirect.com/science/article/pii/S0016508520350216>
112. Finegold SM, Song Y, Liu C, Hecht DW, Summanen P, Könönen E, et al. Clostridium clostridioforme: a mixture of three clinically important species. *Eur J Clin Microbiol Infect Dis Off Publ Eur Soc Clin Microbiol*. 2005;24:319–24.
113. Franzosa EA, Sirota-Madi A, Avila-Pacheco J, Fornelos N, Haiser HJ, Reinker S, et al. Gut microbiome structure and metabolic activity in inflammatory bowel disease. *Nat Microbiol*. 2019;4:293–305.

114. Elmassry MM, Sugihara K, Chankhamjon P, Camacho FR, Wang S, Sugimoto Y, et al. A meta-analysis of the gut microbiome in inflammatory bowel disease patients identifies disease-associated small molecules. *bioRxiv* [Internet]. 2024 [cited 2024 Apr 18];2024.02.07.579278. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10871352/>
115. Desbois AP, Smith VJ. Antibacterial free fatty acids: activities, mechanisms of action and biotechnological potential. *Appl Microbiol Biotechnol*. 2010;85:1629–42.
116. Pham VT, Fehlbaum S, Seifert N, Richard N, Bruins MJ, Sybesma W, et al. Effects of colon-targeted vitamins on the composition and metabolic activity of the human gut microbiome- a pilot study. *Gut Microbes*. 2021;13:1–20.
117. Ning L, Zhou Y-L, Sun H, Zhang Y, Shen C, Wang Z, et al. Microbiome and metabolome features in inflammatory bowel disease via multi-omics integration analyses across cohorts. *Nat Commun* [Internet]. 2023 [cited 2024 Apr 16];14:7135. Available from: <https://www.nature.com/articles/s41467-023-42788-0>
118. Dehne LI, Klemm C, Henseler G, Hermann-Kunz E. The German Food Code and Nutrient Data Base (BLS II.2). *Eur J Epidemiol*. 1999;15:355–9.
119. Nöthlings U, Hoffmann K, Bergmann MM, Boeing H. Fitting portion sizes in a self-administered food frequency questionnaire. *J Nutr*. 2007;137:2781–6.
120. Wang J, Thingholm LB, Skiecevičie J, Rausch P, Kummen M, Hov JR, et al. Genome-wide association analysis identifies variation in Vitamin D receptor and other host factors influencing the gut microbiota. *Nat Genet* [Internet]. 2016;48:1396–406. Available from: <http://dx.doi.org/10.1038/ng.3695>

121. Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience* [Internet]. 2015;4:7. Available from: <http://www.gigasciencejournal.com/content/4/1/7>
122. Wienbrandt L, Ellinghaus D. EagleImp: fast and accurate genome-wide phasing and imputation in a single tool. *Bioinforma Oxf Engl*. 2022;38:4999–5006.
123. Privé F, Aschard H, Ziyatdinov A, Blum MGB. Efficient analysis of large-scale genome-wide data with two R packages: bigstatsr and bigsnpr. *Bioinforma Oxf Engl*. 2018;34:2781–7.
124. Privé F, Arbel J, Aschard H, Vilhjálmsson BJ. Identifying and correcting for misspecifications in GWAS summary statistics and polygenic scores. *HGG Adv*. 2022;3:100136.
125. Thingholm LB, Bang C, Rühlemann MC, Starke A, Sicks F, Kaspari V, et al. Ecology impacts the decrease of Spirochaetes and Prevotella in the fecal gut microbiota of urban humans. *BMC Microbiol* [Internet]. 2021;21:276. Available from: <https://doi.org/10.1186/s12866-021-02337-5>
126. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* [Internet]. 2016;13:581–3. Available from: <http://dx.doi.org/10.1038/nmeth.3869>
127. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* [Internet]. 2007;73:5261–7. Available from: <http://aem.asm.org/cgi/content/abstract/73/16/5261>

128. Lan Y, Wang Q, Cole JR, Rosen GL. Using the RDP classifier to predict taxonomic novelty and reduce the search space for finding novel organisms. *PLoS ONE* [Internet]. 2012;7:e32491. Available from: <http://dx.doi.org/10.1371/journal.pone.0032491>
129. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* [Internet]. 2009;75:7537–41. Available from: <http://aem.asm.org/cgi/content/abstract/AEM.01541-09v1>
130. Price MN, Dehal PS, Arkin AP. FastTree 2 - Approximately maximum-likelihood trees for large alignments. *PLoS ONE* [Internet]. 2010;5:e9490. Available from: <http://dx.doi.org/10.1371/journal.pone.0009490>
131. Douglas GM, Maffei VJ, Zaneveld JR, Yurgel SN, Brown JR, Taylor CM, et al. PICRUSt2 for prediction of metagenome functions. *Nat Biotechnol* [Internet]. 2020;38:685–8. Available from: <https://doi.org/10.1038/s41587-020-0548-6>
132. Webb CO, Ackerly DD, McPeck MA, Donoghue MJ. Phylogenies and community ecology. *Annu Rev Ecol Syst* [Internet]. 2002;33:475–505. Available from: <http://www.annualreviews.org/doi/abs/10.1146/annurev.ecolsys.33.010802.150448>
133. Kembel SW, Cowan PD, Helmus MR, Cornwell WK, Morlon H, Ackerly DD, et al. Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* [Internet]. 2010;26:1463–4. Available from: <http://bioinformatics.oxfordjournals.org/cgi/content/abstract/26/11/1463>
134. Tsirogiannis C, Sandel B. PhyloMeasures: a package for computing phylogenetic biodiversity measures and their statistical moments. *Ecography* [Internet]. 2016;39:709–14. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/ecog.01814>

135. Long JA. jtools: Analysis and Presentation of Social Scientific Data [Internet]. 2022. Available from: <https://cran.r-project.org/package=jtools>
136. Legendre P, Andersson MJ. Distance-based redundancy analysis: Testing multispecies responses in multifactorial ecological experiments. *Ecol Monogr*. 1999;69:1–24.
137. Anderson MJ, Ellingsen KE, McArdle BH. Multivariate dispersion as a measure of beta diversity. *Ecol Lett* [Internet]. 2006;9:683–93. Available from: <http://dx.doi.org/10.1111/j.1461-0248.2006.00926.x>
138. Lahti L, Shetty S. Tools for microbiome analysis in R [Internet]. 2022. Available from: <http://microbiome.github.io/microbiome>
139. Holmes I, Harris K, Quince C. Dirichlet multinomial mixtures: Generative models for microbial metagenomics. *PLoS ONE* [Internet]. 2012;7:e30126. Available from: <http://dx.doi.org/10.1371/journal.pone.0030126>
140. van den Boogaart KG, Tolosana-Delgado R, Bren M. compositions: Compositional Data Analysis [Internet]. 2023. Available from: <https://CRAN.R-project.org/package=compositions>  
<http://www.stat.boogaart.de/compositions/>
141. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* [Internet]. 2014;15:550. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4302049/>
142. Yu G, Wang LG, Han Y, He QY. ClusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS J Integr Biol* [Internet]. 2012;16:284–7. Available from: <https://www.liebertpub.com/doi/abs/10.1089/omi.2011.0118>

143. Kim S. ppcor: An R Package for a Fast Calculation to Semi-partial Correlation Coefficients. *Commun Stat Appl Methods* [Internet]. 2015 [cited 2024 Feb 7];22:665–74. Available from: <http://www.csam.or.kr/journal/view.html?doi=10.5351/CSAM.2015.22.6.665>
144. Brown MB. 400: A Method for Combining Non-Independent, One-Sided Tests of Significance. *Biometrics* [Internet]. 1975 [cited 2024 Feb 7];31:987–92. Available from: <https://www.jstor.org/stable/2529826>
145. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* [Internet]. 2016 [cited 2024 Apr 3];32:2847–9. Available from: <https://doi.org/10.1093/bioinformatics/btw313>
146. Ziyatdinov A, Vázquez-Santiago M, Brunel H, Martinez-Perez A, Aschard H, Soria JM. lme4qtl: linear mixed models with flexible covariance structure for genetic studies of related individuals. *BMC Bioinformatics* [Internet]. 2018 [cited 2023 Apr 3];19:1–5. Available from: <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-018-2057-x>
147. Page L, Brin S. The anatomy of a large-scale hypertextual Web search engine. *Comput Netw*. 1998;30:107–17.
148. Freeman LC. Centrality in social networks conceptual clarification. *Soc Netw*. 1978;1:215–39.
149. Csardi G, Nepusz T. The igraph software package for complex network research. *InterJournal Complex Syst*. 2006;1695.
150. Newman MEJ. Assortative Mixing in Networks. *Phys Rev Lett* [Internet]. 2002;89:208701. Available from: <http://journals.aps.org/prl/abstract/10.1103/PhysRevLett.89.208701>



151. Wasserman S, Faust K. Social Network Analysis: Methods and Applications. Cambridge University Press; 1994.

152. West DB. Introduction to Graph Theory. Prentice Hall; 2001.

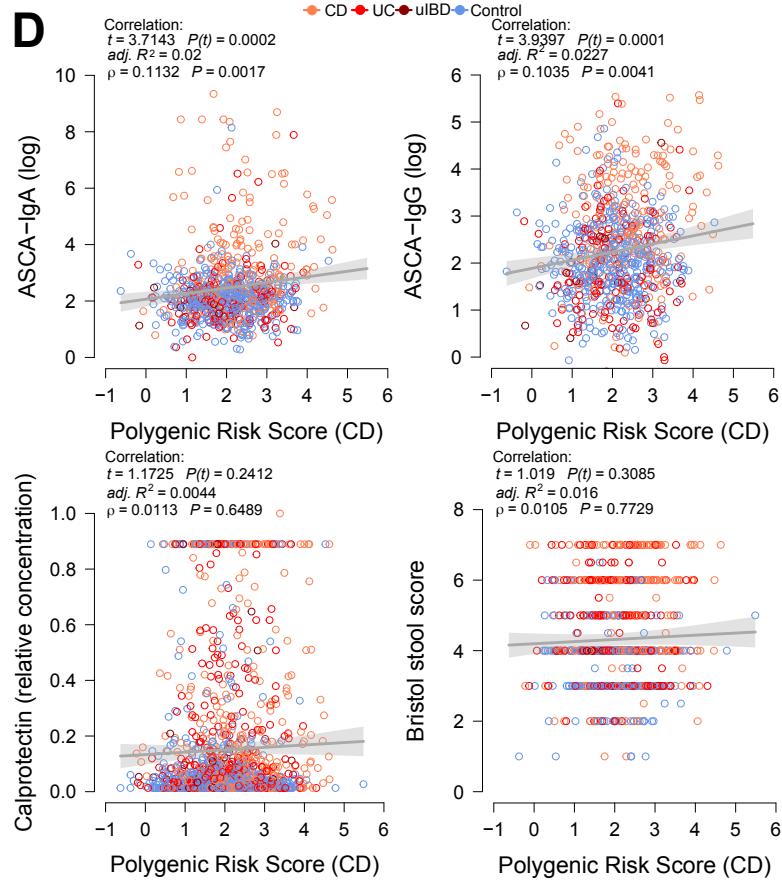
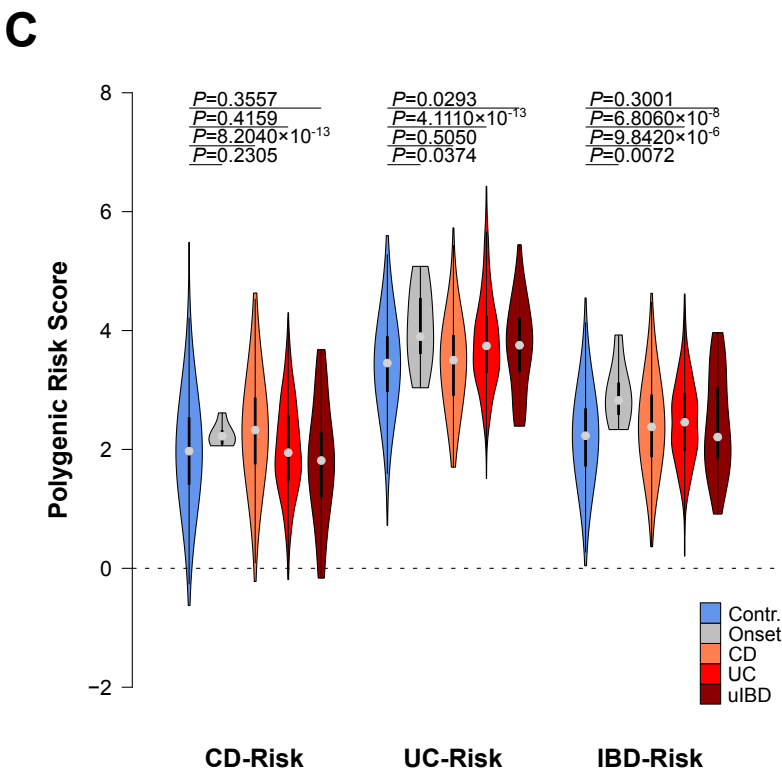
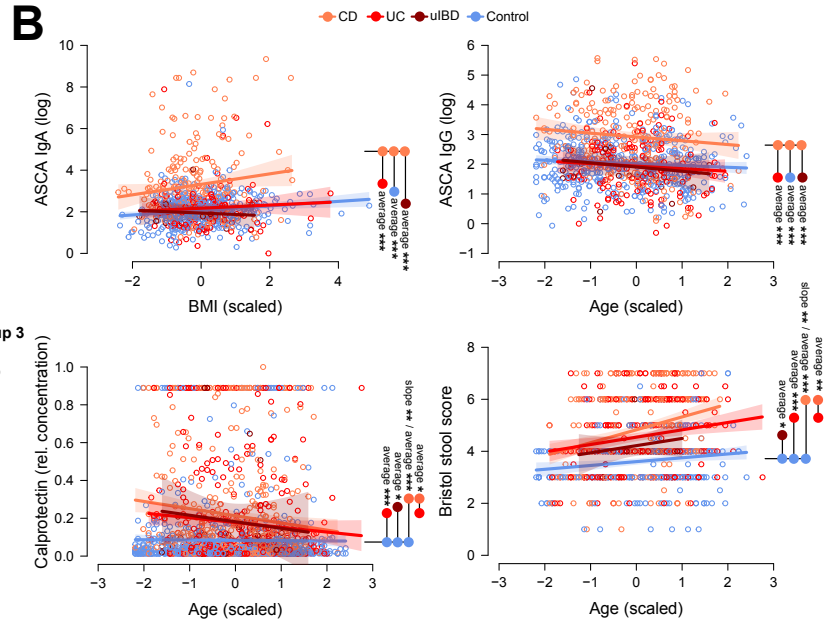
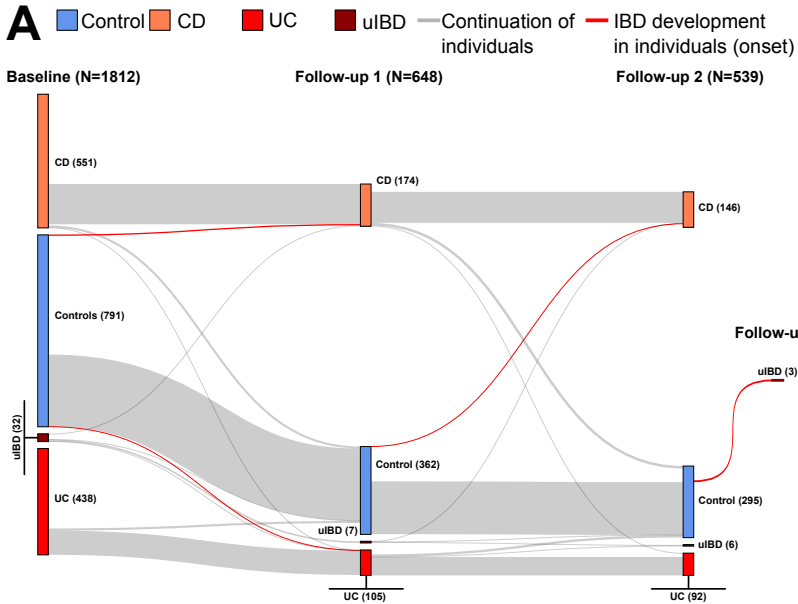
153. Yaveroğlu ÖN, Malod-Dognin N, Davis D, Levnajic Z, Janjic V, Karapandza R, et al. Revealing the Hidden Language of Complex Networks. Sci Rep [Internet]. 2014 [cited 2024 Apr 9];4:4547. Available from: <https://www.nature.com/articles/srep04547>

154. Hočevcar T, Demšar J. Computation of Graphlet Orbits for Nodes and Edges in Sparse Graphs. J Stat Softw [Internet]. 2016 [cited 2023 May 9];71:1–24. Available from: <https://doi.org/10.18637/jss.v071.i10>

155. Venables WN, Ripley BD. The S Language. Mod Appl Stat S [Internet]. New York, NY: Springer New York; 2002. p. 41–68. Available from: [http://dx.doi.org/10.1007/978-0-387-21706-2\\_3](http://dx.doi.org/10.1007/978-0-387-21706-2_3)

156. Zuur AF, Ieno EN, Walker NJ, Saveliev AA, Smith GM. GLM and GAM for Absence–Presence and Proportional Data. Mix Eff Models Ext Ecol R [Internet]. New York, NY: Springer New York; 2009. p. 245–59. Available from: [http://dx.doi.org/10.1007/978-0-387-87458-6\\_10](http://dx.doi.org/10.1007/978-0-387-87458-6_10)

157. Kurtz ZD, Müller CL, Miraldi ER, Littman DR, Blaser MJ, Bonneau RA. Sparse and Compositionally Robust Inference of Microbial Ecological Networks. PLoS Comput Biol [Internet]. 2015;11:e1004226. Available from: <http://dx.doi.org/10.1371/journal.pcbi.1004226>



■ Bacteroidetes ■ Proteobacteria ■ Firmicutes ■ Verrucomicrobia ■ Actinobacteria ■ Fusobacteria ■ Tenericutes ■ Bacteria uncl. ■ Cyanobacteria/Chloroplast ■ Synergistetes ■ Cand. Saccharibacteria ■ Lentisphaerae ■ Spirochaetes ■ Others

