

1 **Mucosal host–microbe interactions associate with clinical phenotypes**
2 **in inflammatory bowel disease**

3 Shixian Hu^{1,2,3*}, Arno R. Bourgonje^{1,*}, Ranko Gacesa^{1,2}, Bernadien H. Jansen¹,
4 Johannes R. Björk¹, Amber Bangma¹, Iwan J. Hidding¹, Hendrik M. van Dullemen¹,
5 Marijn C. Visschedijk¹, Klaas Nico Faber¹, Gerard Dijkstra¹, Hermie J. M. Harmsen⁴,
6 Eleonora A. M. Festen¹, Arnau Vich Vila^{1,2}, Lieke M. Spekhorst^{1,5,#}, Rinse K.
7 Weersma^{1,#}

8 ¹ Department of Gastroenterology and Hepatology, University of Groningen, University
9 Medical Center Groningen, Groningen, the Netherlands

10 ² Department of Genetics, University of Groningen, University Medical Center
11 Groningen, Groningen, the Netherlands

12 ³ Institute of Precision Medicine, the First Affiliated Hospital of Sun Yat-Sen University,
13 Sun Yat-Sen University, Guangzhou, Guangdong, China

14 ⁴ Department of Medical Microbiology, University of Groningen, University Medical
15 Center Groningen, Groningen, the Netherlands

16 ⁵ Department of Gastroenterology and Hepatology, Medisch Spectrum Twente,
17 Enschede, the Netherlands

18 ^{*,#} *These authors contributed equally.*

19 **Correspondence:**

20 Prof. R.K. Weersma, M.D., Ph.D.
21 Department of Gastroenterology and Hepatology
22 University of Groningen, University Medical Center Groningen
23 P.O. Box 30.001, 9700 RB Groningen, the Netherlands
24 E-mail: r.k.weersma@umcg.nl
25 Tel: + 31 50 361 26 20; Fax: +31 50 361 93 06

26 **Abstract**

27

28 Dysregulation of gut mucosal host–microbe interactions is a central feature of
29 inflammatory bowel disease (IBD). To study tissue-specific interactions, we performed
30 transcriptomic (RNA-seq) and microbial (16S-rRNA-seq) profiling of 696 intestinal
31 biopsies derived from 353 patients with IBD and controls. Analysis of transcript-bacteria
32 interactions identified six distinct groups of inflammation-related pathways that were
33 associated with intestinal microbiota, findings we could partially validate in an
34 independent cohort. An increased abundance of *Bifidobacterium* was associated with
35 higher expression of genes involved in fatty acid metabolism, while *Bacteroides* was
36 associated with increased metallothionein signaling. In fibrostenotic Crohn’s disease, a
37 transcriptional network dominated by immunoregulatory genes associated with
38 *Lachnospiraceae* bacteria in non-stenotic tissue. In patients using TNF- α -antagonists,
39 a transcriptional network dominated by fatty acid metabolism genes associated with
40 *Ruminococcaceae*. Mucosal microbiota composition was associated with enrichment of
41 specific intestinal cell types. Overall, we identify multiple host–microbe interactions that
42 may guide microbiota-directed precision medicine.

43

44 **Keywords:** inflammatory bowel disease, gene expression, mucosal microbiota,
45 microbiome, host–microbe interactions.

46 **Main**

47 Inflammatory bowel diseases (IBD), which encompass Crohn's disease (CD) and
48 ulcerative colitis (UC), are chronic inflammatory diseases of the gastrointestinal tract [1].
49 The pathogenesis of IBD is thought to be caused by a complex interplay between
50 inherited and environmental factors, gut microbiota and the host immune system [2,3].

51 Alterations in gut microbiota composition and functionality are commonly observed in
52 patients with IBD, including decreased microbial diversity, decreased abundances of
53 butyrate-producing bacteria and increased proportions of pathobionts [4-8].

54 Interactions between host genetics and the gut microbiome have been studied in both
55 healthy subjects and patients with IBD. For example, we previously focused on host
56 genome–gut microbiota interactions in the context of IBD [9]. However, in order to
57 disentangle disease mechanisms that might underlie the etiology and progression of
58 IBD, there should be a greater focus on mucosal gene expression studies [10].

59 Modulation of host mucosal gene expression by gut microbiota or effects of gene
60 expression on microbial fitness may expose mechanisms that contribute to IBD
61 pathogenesis, knowledge that could be utilized to explore novel therapeutic targets
62 [11,12]. Most studies, however, employ fecal sampling for microbiota characterization,
63 which precludes analysis of local interactions and their immediate impact on host
64 intestinal expression signatures. Such studies examining mucosal gene expression–
65 microbiome associations in the context of IBD previously identified microbial groups
66 associated with host transcripts from immune-mediated and inflammatory pathways [12–
67 15]. In a longitudinal host–microbe interaction study, the chemokine genes *CXCL6* and
68 *CCL20* were negatively associated with the relative abundances of *Eubacterium rectale*
69 and *Streptococcus*, suggesting that these bacteria are more susceptible to the actions
70 of these chemokines [13]. Another study found an inverse association between host
71 expression of *DUOX2*, which produces reactive oxygen species (ROS), and the relative
72 abundance of *Ruminococcaceae*, an association that may suggest ROS-mediated
73 antibacterial effects [16]. However, few studies to date have been able to carry out
74 comprehensive integrated analysis of IBD-associated interaction factors among
75 mucosa-attached microbiota and host intestinal-gene expression.

76 Here we analyzed 696 fresh-frozen intestinal biopsies derived from 337 patients with
77 IBD and 16 non-IBD controls for which we generated both mucosal transcriptomic and
78 microbial characterization using bulk RNA-sequencing and 16S rRNA gene sequencing,
79 respectively. We further combined both datasets to comprehensively investigate mutual
80 mucosal host-microbe interactions and integrated these with the extensive clinical
81 characteristics collected. Following this approach, we aimed to investigate mucosal
82 host–microbe interactions while disentangling disease-, location- and inflammation-
83 specific associations (a graphical representation of the study workflow is presented in
84 **Figure 1**). Most importantly, we could study the associations between mucosal host–
85 microbe interactions and clinical phenotypes of patients with IBD. Finally, we also
86 sought to replicate our main results in data from a smaller independent, publicly
87 available cohort [13].

88 **Results**

89 **Cohort description**

90 Demographic and clinical characteristics of the study population are presented in **Table**

91 **1.** In total, we included 640 intestinal biopsies from 337 patients with IBD and 56

92 intestinal biopsies from 16 non-IBD controls. Biopsies were derived from the colon

93 (64.4%) and ileum (35.6%), and patients with CD and UC were equally represented

94 among inflamed (CD: 53.8%, UC: 46.2%) and non-inflamed (CD: 55.4%, UC: 44.6%)

95 biopsies. Mean age and the proportion of smokers were higher among controls ($P<0.01$

96 and $P=0.01$, respectively). Among biopsies derived from patients with IBD, the

97 proportion of steroid users was higher among patients from whom inflamed biopsies

98 were collected ($P<0.01$). Remaining patient characteristics were evenly distributed

99 among groups without significant differences.

100 **Table 1.** Demographic and clinical characteristics of the study population compared

101 between the inflamed and non-inflamed dataset.

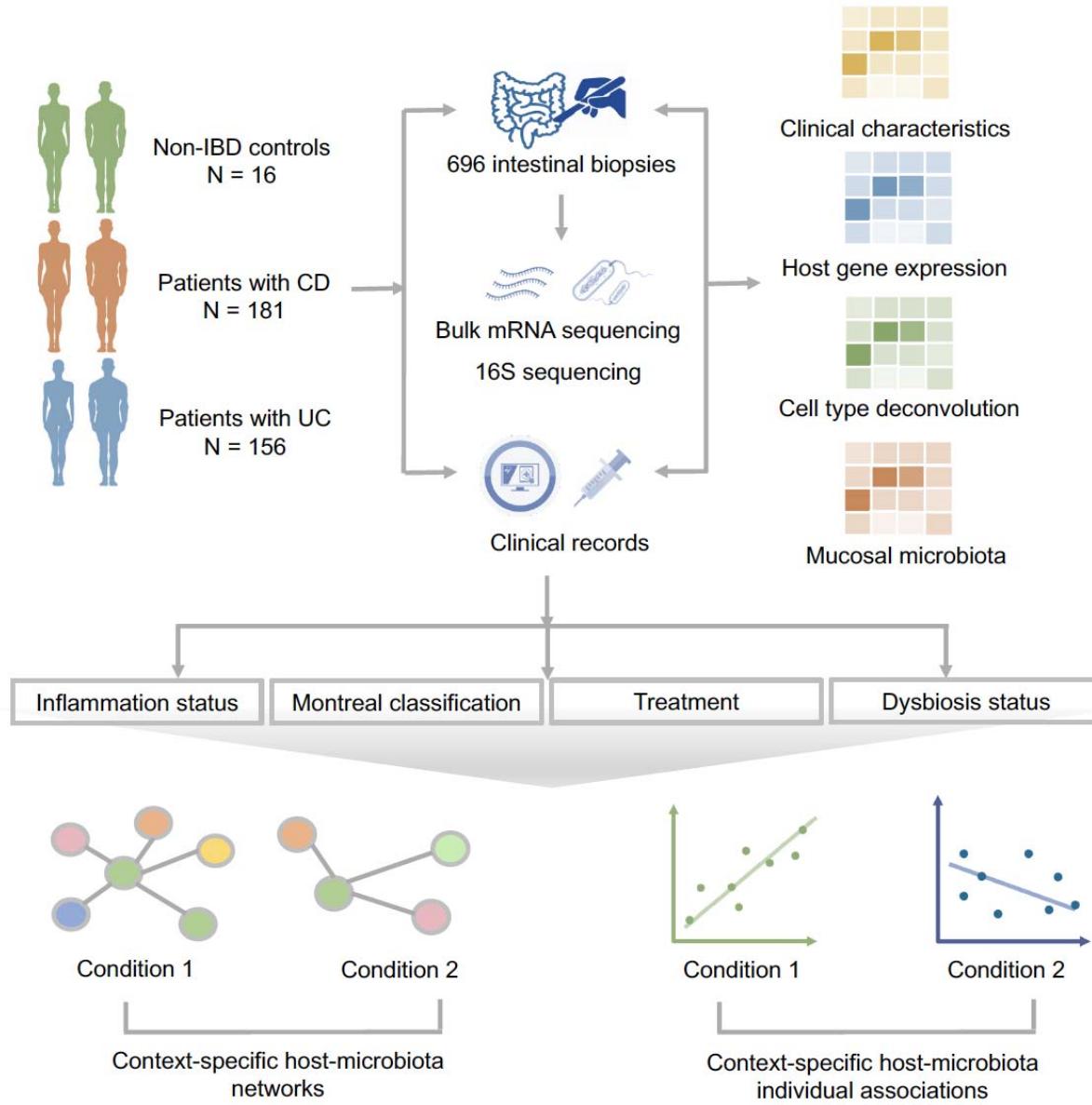
Variable	Total	IBD		Non-IBD	P-value
		Inflamed biopsies	Non-inflamed biopsies		
	<i>n</i> = 696	<i>n</i> = 212	<i>n</i> = 428	<i>n</i> = 56	
Biopsy inflammation, <i>n</i> (%)					
<i>Inflamed</i>	212 (30.5)	212 (100)	-	-	
<i>Non-inflamed</i>	428 (61.5)	-	428 (100)	-	
Biopsy location, <i>n</i> (%)					<0.01
<i>Ileum</i>	248 (35.6)	66 (31.1)	173 (40.4)	9 (16.1)	
<i>Colon</i>	448 (64.4)	146 (68.9)	255 (59.6)	47 (83.9)	
Diagnosis or control, <i>n</i> (%)					0.74

<i>CD</i>	351 (50.4)	114 (53.8)	237 (55.4)	-	
<i>UC</i>	289 (41.5)	98 (46.2)	191 (44.6)	-	
<i>Controls</i>	56 (8.0)	-	-	56 (100)	
Age at biopsy (years)	43.1 ± 15.3	43.0 ± 15.8	42.3 ± 15.5	45.5 ± 10.8	<0.01
Sex, <i>n</i> (%)					
<i>Male</i>	322 (46.3)	92 (43.4)	188 (43.9)	42 (75.0)	
<i>Female</i>	374 (53.7)	120 (56.6)	240 (56.1)	14 (25.0)	
BMI (kg/m ²)	25.7 ± 4.5	25.7 ± 4.6	25.7 ± 4.7	24.7 ± 2.5	
Current smoking, <i>n</i> (%)					0.01
<i>Yes</i>	145 (20.8)	37 (17.5)	88 (20.6)	20 (35.7)	
<i>No</i>	551 (79.2)	175 (82.5)	340 (79.4)	36 (64.3)	
Montreal classification					
Montreal Age (A), <i>n</i> (%)	638 (99.7)	212 (100)	426 (99.5)	-	0.95
A1 (≤16 years)	73 (11.4)	23 (10.8)	50 (11.7)	-	
A2 (17–40 years)	397 (62.0)	132 (62.3)	265 (62.2)	-	
A3 (>40 years)	168 (26.3)	57 (26.9)	111 (26.1)	-	
Montreal Location (L), <i>n</i> (%)	333 (94.9)	108 (94.7)	225 (94.9)	-	0.83
L1 (ileal disease)	66 (19.8)	18 (16.7)	48 (21.3)	-	
L2 (colonic disease)	51 (15.3)	17 (15.7)	34 (15.1)	-	
L3 (ileocolonic disease)	172 (51.7)	58 (53.7)	114 (50.7)	-	
L1 + L4	12 (3.6)	3 (2.8)	9 (4.0)	-	
L2 + L4	6 (1.8)	3 (2.8)	3 (1.3)	-	
L3 + L4	26 (7.8)	9 (8.3)	17 (7.6)	-	

Montreal Behavior (B), <i>n</i> (%)	333 (94.9)	108 (94.7)	225 (94.9)	-	0.31
B1 (non-stricturing, non-penetrating)	146 (43.8)	53 (49.1)	93 (41.3)	-	
B2 (stricturing)	59 (17.7)	20 (18.5)	39 (17.3)	-	
B3 (penetrating)	30 (9.0)	9 (8.3)	21 (9.3)	-	
B1 + P (perianal disease)	35 (10.5)	10 (9.3)	25 (11.1)	-	
B2 + P (perianal disease)	48 (14.4)	15 (13.9)	33 (14.7)	-	
B3 + P (perianal disease)	15 (4.5)	1 (0.9)	14 (6.2)	-	
Montreal Extension (E), <i>n</i> (%)	246 (85.1)	83 (84.7)	163 (85.3)	-	0.95
E1 (proctitis)	19 (7.7)	7 (8.4)	12 (7.4)	-	
E2 (left-sided colitis)	75 (30.5)	25 (30.1)	50 (30.7)	-	
E3 (pancolitis)	152 (61.8)	51 (61.4)	101 (62.0)	-	
Montreal Severity (S), <i>n</i> (%)	207 (71.6)	68 (69.4)	139 (72.8)		0.70
S0 (remission)	11 (5.3)	3 (4.4)	8 (5.8)		
S1 (mild)	28 (13.5)	9 (13.2)	19 (13.7)		
S2 (moderate)	109 (52.7)	33 (48.5)	76 (54.7)		
S3 (severe)	59 (28.5)	23 (33.8)	36 (25.9)		
Medication use					
Aminosalicylates, <i>n</i> (%)	271 (42.3)	91 (42.9)	180 (42.1)	-	0.87
Thiopurines, <i>n</i> (%)	210 (32.8)	66 (31.1)	144 (33.6)	-	0.53
Steroids, <i>n</i> (%)	262 (40.9)	105 (49.5)	157 (36.7)	-	<0.01
Methotrexate, <i>n</i> (%)	44 (6.9)	18 (8.5)	26 (6.1)	-	0.32

TNF- α -antagonists, <i>n</i> (%) [†]	113 (17.7)	35 (16.5)	78 (18.4)	-	0.58
Clinical disease activity					
HBI	324 (92.3)	104 (91.2)	220 (92.8)	-	0.18
Remission (<5)	205 (63.3)	60 (57.7)	145 (65.9)	-	
Active disease (≥ 5)	119 (36.7)	44 (42.3)	75 (34.1)	-	
SCCAI	257 (88.9)	84 (85.7)	173 (90.6)	-	0.14
Remission (≤ 2)	152 (59.1)	44 (52.4)	108 (62.4)	-	
Active disease (> 2)	105 (40.9)	40 (47.6)	65 (37.6)	-	
Surgical history					
Ileocecal resection, <i>n</i> (%)	132 (20.6)	40 (18.9)	92 (21.5)	-	0.47
Colon resection (or partial), <i>n</i> (%)	146 (22.8)	55 (25.9)	91 (21.3)	-	0.19
Small intestinal (partial) resection, <i>n</i> (%)	82 (12.8)	29 (13.7)	53 (12.4)	-	0.71

102 Data are presented as proportions *n* with corresponding percentages (%), mean \pm standard deviation
103 (SD) or as median [interquartile range, IQR] in case of continuous variables. *P*-values ≤ 0.05 were
104 considered statistically significant. [†]Use of TNF- α -antagonists included use of infliximab, adalimumab,
105 golimumab and certolizumab pegol. Abbreviations: BMI, body-mass index; CD, Crohn's disease; HBI,
106 Harvey-Bradshaw Index; IBD, inflammatory bowel disease; TNF- α , tumor necrosis factor alpha; SCCAI,
107 Simple Clinical Colitis Activity Index; UC, ulcerative colitis.



108

109 **Figure 1. Methodological workflow of the study.** The study cohort consisted of 337 patients with IBD
110 (CD: n=181, UC: n=156) and 16 non-IBD controls, from whom 696 intestinal biopsies were collected (IBD:
111 n=640, controls: n=56) and processed to perform bulk mucosal mRNA-sequencing and 16S gene rRNA
112 sequencing. Detailed phenotypic data were extracted from clinical records for all study participants. In
113 total, 251 ileal biopsies (CD: n=186, UC: n=56, controls: n=9) and 445 colonic biopsies (CD: n=165, UC:
114 n=233, controls: n=47) were included: 212 biopsies derived from inflamed regions and 484 from non-
115 inflamed regions. Mucosal gene expression and bacterial abundances were systematically analyzed in
116 relation to different (clinical) phenotypes: presence of tissue inflammation, Montreal disease classification,

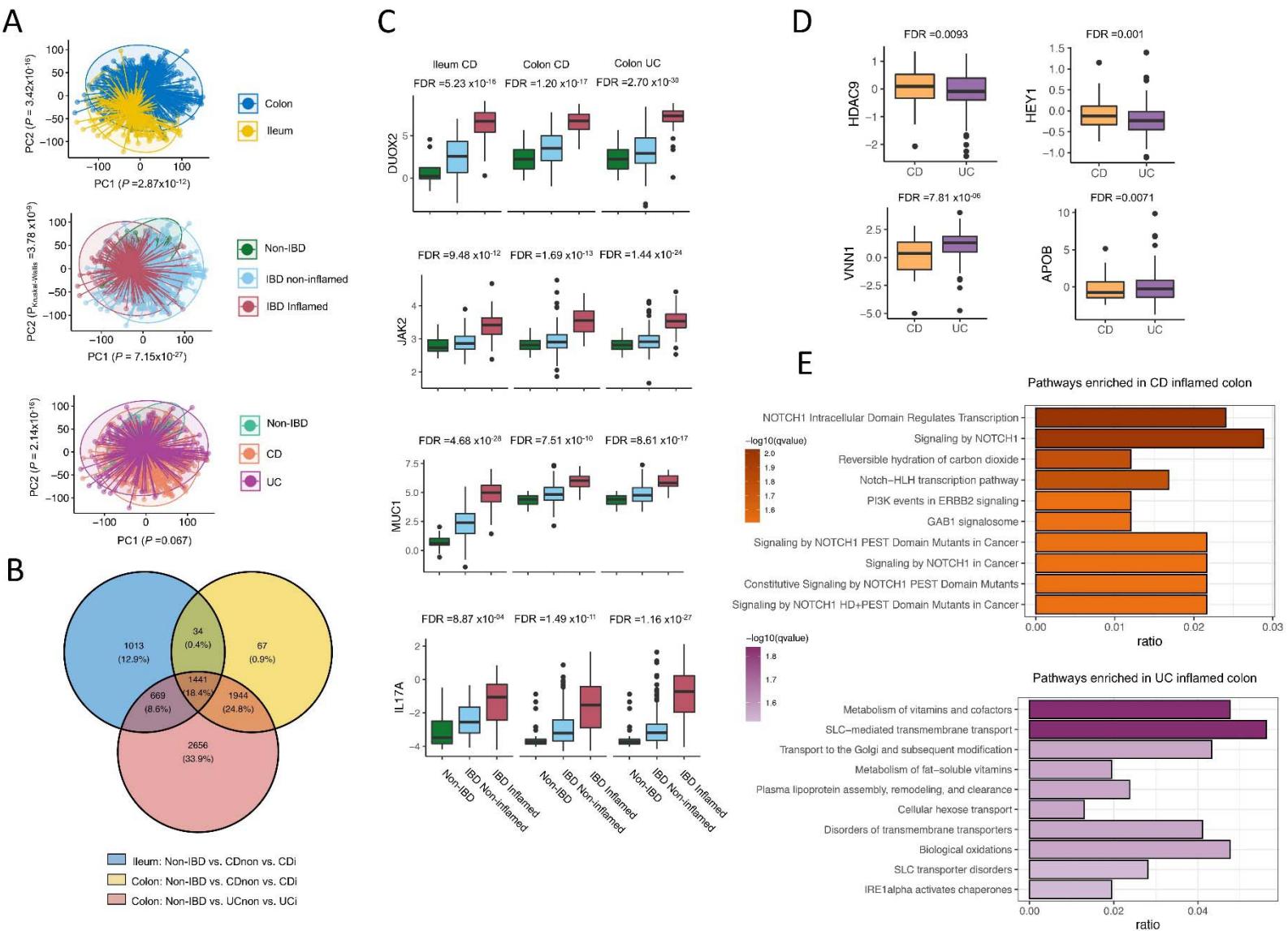
117 medication use (e.g. TNF- α -antagonists) and dysbiotic status. Pathway-based clustering and network
118 analysis (Sparse-CCA and centrLCC analysis) and individual pairwise gene–taxa associations were
119 investigated to identify host–microbiota interactions in different contexts. We then analyzed the degree to
120 which mucosal microbiota could explain the variation in intestinal cell type–enrichment (estimated by
121 deconvolution of bulk RNA-seq data). To confirm our main findings, we used publicly available mucosal
122 16S and RNA-seq datasets for external validation [13].

123

124 **Mucosal gene expression reflects tissue specificity, inflammatory status and**
125 **disease subtypes**

126 Principal component analysis (PCA) showed that gene transcriptional patterns could be
127 stratified by biopsy location (ileum vs. colon), inflammatory status (non-inflamed vs.
128 inflamed) and IBD subtype (CD vs. UC) in the first two components (**Fig. 2A**), consistent
129 with previous observations [13]. Tissue location and inflammatory status were
130 significantly associated with the first two PCs (biopsy location, ileum vs. colon:
131 $P_{\text{Wilcoxon}}=2.87\times10^{-12}$; biopsy inflammatory status, $P=7.15\times10^{-27}$), whereas disease/control
132 status (CD vs. UC vs. controls) was associated with the second PC ($P=2.14\times10^{-16}$).

133 Inflammation-associated gene expression showed overlap between inflamed biopsies
134 from ileal CD, colonic CD and UC (**Fig. 2B**). Differential expression analyses between
135 non-IBD controls, non-inflamed and inflamed biopsies in all these three groups revealed
136 3157, 3486, and 6710 differentially expressed genes (DEGs), respectively (FDR<0.05)
137 (**Supplementary Table S1**). These DEGs fall mainly within interleukin signaling,
138 neutrophil degranulation and extracellular matrix (ECM) organization pathways
139 (FDR_{Fisher}<0.05, **Extended Data Fig. S1**). Overlapping results from all three differential
140 expression analyses identified 1437 shared DEGs, including *DUOX2*, *MUC1*, *JAK2*,
141 *OSM* and *IL17A* (**Fig. 2C**). We also observed an enrichment of these DEGs in IBD-
142 associated genomic loci ($P_{\text{Fisher}}=9.6\times10^{-9}$) [2].



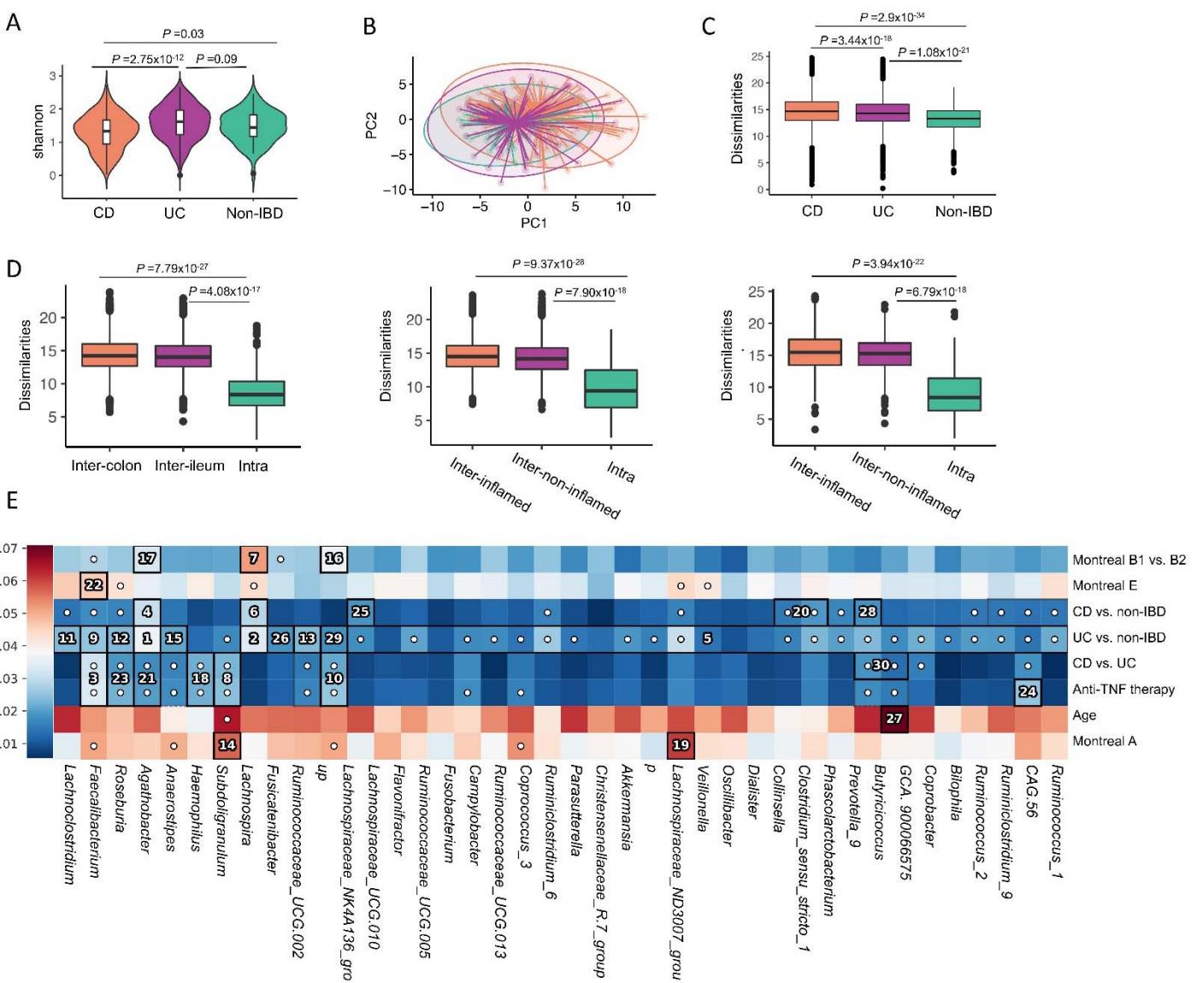
144 **Figure 2. Mucosal host gene expression patterns in intestinal tissue from patients with IBD and controls.** **a**, Principal component
145 analysis, labeled by tissue location (ileum/colon), inflammatory status (non-inflamed/inflamed) and disease diagnosis (control/CD/UC),
146 shows that variation in host gene expression can be significantly explained by tissue location and inflammatory status. **b**, Venn diagram of
147 inflammation-associated genes from three comparisons: 1) ileal tissue from controls vs. non-inflamed tissue from patients with CD vs.
148 inflamed tissue from patients with CD, 2) colonic tissue from controls vs. non-inflamed tissue from patients with CD vs. inflamed tissue
149 from patients with CD and 3) colonic tissue from controls vs. non-inflamed tissue from patients with UC vs. inflamed tissue from patients
150 with UC (all FDR <0.05). **c**, Relevant examples of four inflammation-associated genes, *DUOX2*, *JAK2*, *MUC1* and *IL17A*, illustrating the
151 presence of tissue inflammation (FDR <0.05). **d**, Relevant examples of inflammation-associated genes differentially expressed between
152 patients with CD and UC (keeping tissue location and inflammatory status constant) showing higher expression of *HDAC9* (histone
153 deacetylase 9) and *HEY1* (hairy/enhancer-of-split related with YRPW motif protein 1) in patients with CD and higher expression of *VNN1*
154 (pantetheinase) and *APOB* (apolipoprotein B) in patients with UC. **e**, Analysis of pathways associated with either the presence of CD
155 (orange) or UC (purple) demonstrates that genes upregulated in CD are mainly associated with Notch-1 signaling, whereas pathways
156 upregulated in UC are mainly related to vitamin and cofactor metabolism, SLC-mediated transmembrane transport and intracellular protein
157 modification. Pathways were annotated using the Reactome pathway database. CDi, inflamed tissue from patients with Crohn's disease.
158 CD-non, non-inflamed tissue from patients with Crohn's disease. FDR, false discovery rate. PC, principal component. UCi, inflamed tissue
159 from patients with ulcerative colitis. UC-non, non-inflamed tissue from patients with ulcerative colitis.

160 We then investigated the genes differentially expressed between inflamed colonic tissue
161 from patients with CD and UC. In total, 1466 genes were differentially abundant, of
162 which 733 (50%) were overrepresented in CD and 733 (50%) in UC (FDR<0.05)
163 (**Supplementary Table S2**). Pathway enrichment analysis showed the Notch-1
164 signaling pathway (e.g. *HDAC9* and *HEY1*, **Fig. 2D**) to be highly upregulated in CD
165 compared to UC, whereas vitamin, cofactor and lipoprotein metabolism pathways (e.g.
166 *VNN1* and *APOB*, **Fig. 2D**) were more pronounced in UC (**Fig. 2E**), which corroborates
167 previous findings [17-20]. Cell type–deconvolution revealed that plasma cells,
168 endothelial cells and Th2-lymphocytes were significantly increased in UC compared
169 with CD (FDR<0.05, **Supplementary Table S3**), suggesting that distinct immunological
170 mechanisms are involved in CD and UC.

171

172 **Mucosal microbiota composition is highly personalized**

173 The most common bacterial phylum observed across all tissue samples was
174 Bacteroidetes (CD: 58%, UC: 58%, controls: 66%), followed by Firmicutes (CD: 27%,
175 UC: 33%, controls: 23%) and Proteobacteria (CD: 14%, UC: 8%, controls: 9%).
176 Interestingly, the overall mucosa-attached microbial composition was similar between
177 colonic and ileal biopsies and independent of inflammation (**Extended Data Fig. S2**).
178 Only seven bacterial taxa were differentially abundant between patients and controls
179 (**Supplementary Tables S4-5**), consistent with previous findings [13,21,22].
180 Shannon diversity was significantly lower in samples from patients with CD compared to
181 UC and non-IBD controls ($P=2.75\times10^{-16}$ and $P=0.03$, respectively, **Fig. 3A**). This
182 difference was still present when comparing only colonic biopsies from patients with CD
183 to those from UC, indicating that this difference was not solely attributable to ileal CD
184 (**Extended Data Fig. S3**). Differences in microbial communities between tissue samples
185 were evaluated by quantifying the Aitchison's distance (**Fig. 3B-C**). We obtained
186 comparable findings when we externally validated our results using data derived from
187 the HMP2 cohort (**Extended Data Fig. S4**) [13].



189 **Figure 3. Overall characterization of mucosa-attached microbiota in patients with IBD and controls.** **a**, Microbial alpha-diversity (Shannon
190 index) was lowest in patients with CD (n=351) compared to patients with UC (n=289) and non-IBD controls (n=56). **b**, PCA plot based on
191 Aitchison's distances demonstrates the microbial dissimilarity of the mucosa-attached microbiota (colors as in **a**). **c**, Microbial dissimilarity
192 (Aitchison's distances) comparison between non-IBD control, CD and UC. Microbial dissimilarity is highest in biopsies from patients with CD,
193 followed by patients with UC and non-IBD controls. **d**, Microbial dissimilarity is higher in samples from different individuals when compared to
194 paired samples from the same individual, which includes paired inflamed–non-inflamed tissue from ileum and colon (left panel, inter-colon:
195 n=11,430, inter-ileum: n=7,377, intra: n=203), paired colonic tissue samples from inflamed and non-inflamed areas (middle panel, inter-inflamed:
196 n=7,372, inter-non-inflamed: n=8,369, intra: n=166) and paired ileal tissue samples from inflamed and non-inflamed areas (right panel, inter-
197 inflamed: n=1,590, inter-non-inflamed: n=1,592, intra: n=73). **e**, Hierarchical analysis performed using an end-to-end statistical algorithm (HALIA)
198 indicates the main phenotypic factors that correlate with intestinal mucosal microbiota composition. Heatmap color palette indicates normalized
199 mutual information. Numbers or dots in cells identify significant pairs of features (phenotypic factors vs. bacterial taxa) in patients with IBD and
200 controls. CD, Crohn's disease. PCA, principal coordinate analysis. UC, ulcerative colitis.

201 Intra-individual microbial dissimilarity was lowest in all our comparative analyses of
202 paired tissue samples (**Fig. 3D**). Hierarchical clustering analysis performed on paired
203 samples demonstrated a clear tendency of these samples to cluster together, a finding
204 that we could also replicate in the HMP2 cohort data (**Extended Data Fig. S5A**) [13].
205 Overall, our data demonstrate that the composition of the mucosal microbiota is highly
206 personalized and that inter-individual variability dominates over the effects of tissue
207 location or inflammatory status.

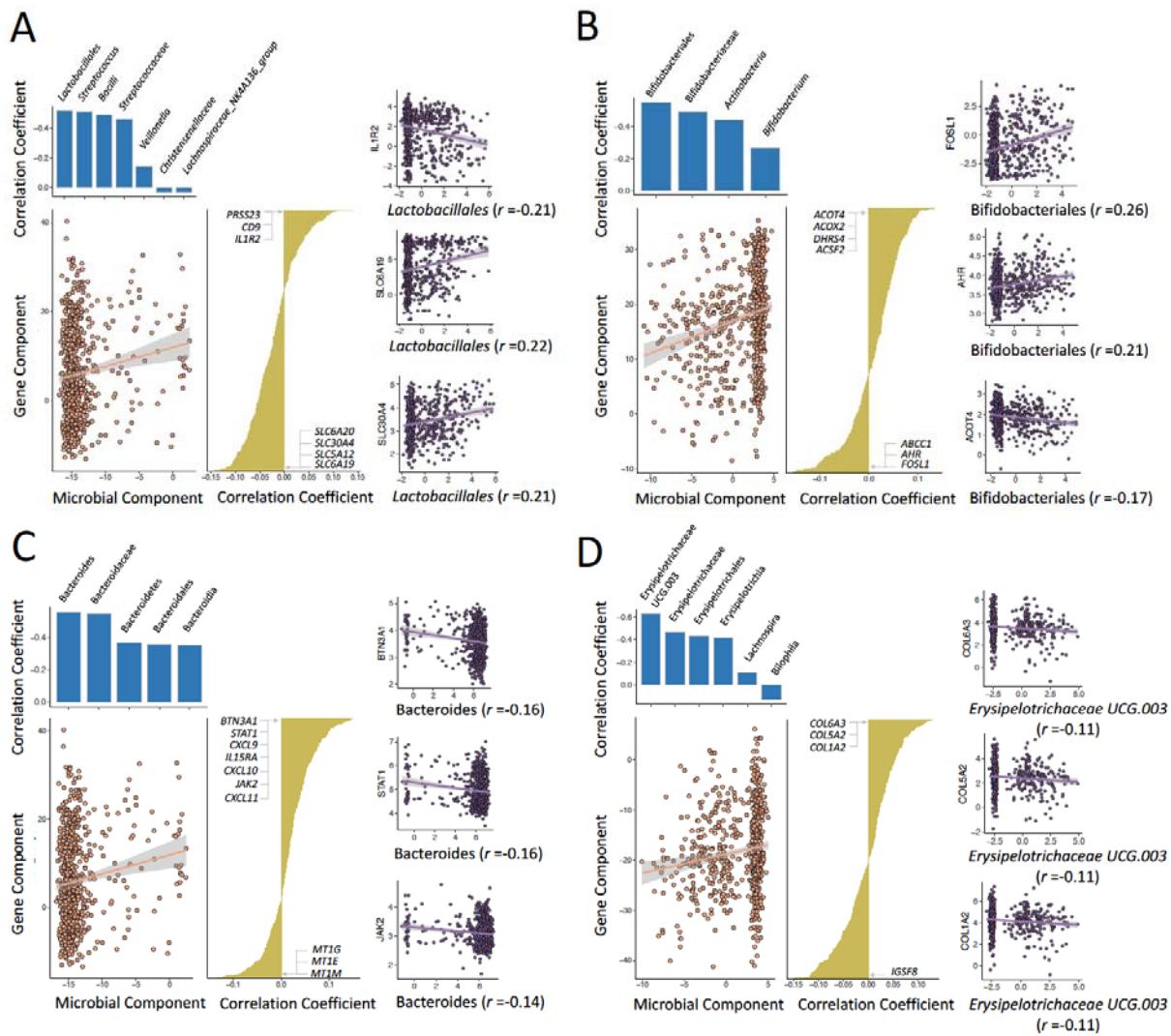
208 We then aimed to identify phenotypic factors that shape the composition of the mucosal
209 microbiota using Hierarchical All-against-All association (HAllA) analysis. This allowed
210 us to study the relative associations between microbial taxa and phenotypic factors and
211 disease characteristics (**Fig. 3E, Supplementary Table S6**). Analysis at bacterial genus
212 level revealed that the main factors correlating with mucosal microbiota composition are
213 stricturing disease in CD (fibrostenotic CD, Montreal B2), usage of TNF- α -antagonists,
214 age at time of sampling, age of onset and the comparisons of patients with CD vs.
215 controls, UC vs. controls and CD vs. UC. In contrast, inflammatory status and tissue
216 location did not show a significant effect, and this was also the case within the HMP2
217 cohort data (**Extended Data Fig. S5B**). These findings are in line with several previous
218 observations from which age at diagnosis, age at sampling and TNF- α -antagonist use
219 emerged as critical determinants of mucosal microbiota composition [22].

220

221 **Distinct host–microbe interaction modules are identified in relation to IBD**

222 To capture the main microbial taxa associated with inflammation-associated gene
223 expression, we combined the data and performed sparse canonical correlation analysis
224 (sparse-CCA) on 1437 inflammation-associated genes and 131 microbial taxa. This
225 approach enabled us to identify gene pathways and groups of microbiota and their
226 potential correlations. In total, we found six distinct pairings of groups of genes with
227 bacterial taxa to be significantly correlated with each other (FDR<0.05, **Supplementary**
228 **Tables S7–S18**). To prioritize the individual genes and bacteria involved in the sparse-
229 CCA analysis, we performed individual pairwise gene–bacteria associations, which
230 revealed 312 significant gene–bacteria pairs, with most pairs (94.17%) overlapping with

231 the sparse-CCA results. We then replicated these associations in the HMP2 cohort
232 (Spearman correlation $\rho=0.16$, $P=0.005$, **Supplementary Table S19, Extended Data**
233 **Fig. S6, Methods**). Further details on the most intriguing individual pairwise gene–
234 bacteria associations are discussed in **Box 1**.



235

236 **Figure 4. Mucosal host–microbe interaction modules in the context of IBD.** Sparse canonical
237 correlation analysis (sparse-CCA) was performed to identify distinct correlation modules of mucosal gene
238 expression vs. mucosal microbiota through the identification of sparse linear combinations of two
239 separate distance matrices that are highly correlated. Using 1437 inflammation-related genes and 131
240 microbial taxa as input, we identified six distinct pairs of significantly correlated gene–microbe
241 components (FDR<0.05). **a**, A diverse group of mainly lactic acid producing bacteria (LAB) represented
242 by order Lactobacillales, genus *Streptococcus*, class Bacilli, family Streptococcaceae and, to a lesser

243 extent, genus *Veillonella*, family *Christensenellaceae* and the *Lachnospiraceae* NK4A136 group is
244 associated with host pathways predominantly related to solute transport and liver metabolism. **b**, The
245 abundance of mucosal *Bifidobacterium* bacteria is inversely associated with host fatty acid metabolism
246 pathways (e.g. *ACOT4*, *ACOX2* and *ACSF2*) and positively associated with expression of specific genes,
247 including *FOSL1*, *AHR* and *ABCC1/MRP1*. **c**, Mucosal *Bacteroides* bacteria inversely correlate with
248 expression of genes representing host interleukin signaling pathways (e.g. *STAT3*, *JAK2*, *CXCL9* and
249 *IL15RA*) but positively correlate with expression of genes representing metal ion response and
250 metallothionein pathways (e.g. the metal ion response transcription factors *MT1G*, *MT1E* and *MT1M*). **d**,
251 Mucosal *Erysipelotrichaceae* abundance inversely associates with expression of genes involved in
252 collagen biosynthesis and collagen trimerization (e.g. *COL1A2*, *COL4A1* and *COL5A2*). Details of the two
253 other significantly correlated pairs of components are presented in **Box 2**.

254 *Mucosal lactic acid-producing bacteria positively correlate with nutrient uptake and*
255 *solute transport*

256 In the first significant pair of correlated components (component pair 1, $P=5.72\times 10^{-14}$,
257 FDR<0.05), the bacterial component is represented by bacteria from order
258 Lactobacillales, family *Streptococcaceae*, class Bacilli and genus *Streptococcus* and, to
259 a lesser extent, genus *Veillonella*, family *Christensenellaceae* and the *Lachnospiraceae*
260 NK4A136 group (**Supplementary Tables S7-S8**). This bacterial component is mainly
261 represented by lactic acid producing bacteria (LABs, including Lactobacillales, Bacilli,
262 *Streptococcaceae*, *Streptococcus*) that actively participate in physiological food
263 digestion, particularly carbohydrate fermentation, with lactic acid being their main
264 metabolic product. Many of these bacterial groups are associated with genes involved in
265 pathways related to solute transport and liver metabolism, including SLC-mediated
266 transmembrane transport of bile salts, organic acids, metal ions and amine compounds;
267 amino acid transport; biological oxidation; cytochrome P450 enzymes and the ephrin
268 signaling pathway (involved in the migration of intestinal epithelial cells along the crypt-
269 villus axis).

270 LABs are widely present in commercially available probiotics, and their beneficial effects
271 on intestinal epithelial health are well-recognized [23]. SLC transporters mediate the
272 bidirectional passage of nutrients such as sugars, amino acids, vitamins, electrolytes
273 and drugs across the intestinal epithelium [24]. Although SLC transporters are often
274 found to be dysregulated in patients with IBD (particularly CD), their expression may be

275 stimulated and subsequently restored by commensal probiotic bacteria [25-27]. Taken
276 together, however, we foresee that this host–microbe interaction component might not
277 be IBD-specific as the genes and bacteria involved have important physiological
278 functions in nutrient digestion and absorption.

279 *Mucosa-residing Bifidobacterium species show significant interplay with host fatty acid*
280 *metabolism and bile acid transport pathways*

281 The second pair of significantly associated components (component pair 3, $P=1.89\times10^{-8}$,
282 FDR<0.05) is predominantly represented by bifidobacteria (**Supplementary Tables**
283 **S9-S10**). The top associated pathways are represented by genes involved in fatty acid
284 metabolism, including fatty acid biosynthesis (e.g. *ACOT4* and *ACSF2*), arachidonic
285 acid metabolism (e.g. *CYP2J2* and *EPHX2*) and genes involved in peroxisomal protein
286 import and fatty acid synthesis (e.g. *PEX5* and *ACOT4*), and these genes are all
287 inversely associated with the bacterial component. In contrast, the genes *AHR*
288 (encoding for the aryl hydrocarbon receptor) and *ABCC1* (encoding multidrug resistance
289 protein 1) are positively correlated with the bacterial component. The inverse
290 associations between bifidobacteria and the expression of genes involved in
291 adipogenesis are consistent with findings from animal and small-scale human studies
292 that investigated the effects of treatment with *Bifidobacterium* species on fatty acid
293 metabolism [28-32]. Our findings may reflect the anti-inflammatory and anti-lipogenic
294 role of bifidobacteria, which has previously been demonstrated in experimental settings,
295 and may support the therapeutic potential of microbiome-directed interventions in
296 attenuating or preventing colitis [31,32].

297 *Mucosal Bacteroides associate with host interleukin signaling and metal ion response*
298 *pathways*

299 The third pair of significantly correlated components (component pair 7, $P=1.28\times10^{-4}$,
300 FDR<0.05) is represented by Bacteroidetes. Twenty-four different pathways were
301 significantly associated with this microbial component (**Supplementary Tables S11-**
302 **S12**). A number of interferon signaling pathways (e.g. IFN- α , IFN- β and IFN- γ as well as
303 the IL-2, IL-4, IL-6, IL-10, IL-12 and IL-13 signaling pathways) are all inversely
304 associated with the microbial component. In addition, metal ion response and

305 metallothionein pathways (e.g. metal ion transcription factors *MT1A*, *MT1E*, *MT1F*,
306 *MT1G* and others) are positively associated with the microbial component. Taken
307 together, these observations could suggest a predominance of potentially beneficial
308 *Bacteroides* species associated with this component. Previous studies have shown that
309 *Bacteroides* can exert either beneficial, mutualistic, or pathogenic effects on the host,
310 depending on local interactions, intestinal location and nutrient availability [21,33]. The
311 co-occurrence of *Bacteroides* with lower expression of interleukin signaling pathways is
312 supported by experimental work that found potential anti-inflammatory and protective
313 roles for these bacteria in the context of intestinal inflammation [34,39,40]. Still, the
314 relative contributions of each of these species, as well as their behavior in the context of
315 intestinal inflammation, remains elusive, although our data might reflect an
316 overrepresentation of anti-inflammatory members [34-39,43]. The positive associations
317 between *Bacteroides* and expression of metal ion response genes and metallothioneins
318 (MTs) are intriguing in the context of IBD because aberrant MT homeostasis and
319 intracellular zinc metabolism have been implicated in disease pathophysiology [44-48].

320 *Mucosal Erysipelotrichaceae bacteria interact with collagen biosynthesis pathways*
321 In the fourth pair of significantly correlated components (component pair 8, $P=1.22\times 10^{-4}$,
322 FDR<0.05), the microbial component, represented by the family *Erysipelotrichaceae*, is
323 inversely associated with the expression of genes belonging to a wide range of ECM
324 and collagen genes that are involved in collagen biosynthesis, integrin cell surface
325 interactions, collagen chain trimerization, collagen fibril cross-linking, collagen fibril
326 assembly, ECM proteoglycans, collagen degradation and related pathways
327 (**Supplementary Tables S13-S14**). Similar to *Bacteroides*, the precise role of
328 *Erysipelotrichaceae* in the context of IBD has not yet been fully elucidated. Some studies
329 found lower abundances of *Erysipelotrichaceae* in patients with new-onset CD [49] and
330 postoperative active CD [50], whereas others reported higher levels of
331 *Erysipelotrichaceae* in the context of ileitis [51] and TNF-regulated CD-like transmural
332 inflammation [52]. These inconsistencies have been suggested to be due to
333 *Erysipelotrichaceae* behaving differently in response to intestinal inflammation, but they
334 may also reflect incomplete characterization of the precise species that belong to the

335 family of *Erysipelotrichaceae* [53]. Interactions between *Erysipelotrichaceae* and
336 ECM/collagen remodeling pathways have not yet been reported in the context of IBD,
337 but they would be particularly relevant because fibrosis occurs in a large fraction of
338 patients with CD and *Erysipelotrichaceae* bacteria have been associated with fibrotic
339 conditions beyond IBD [54-60].

340

341 **Patients with fibrostenotic CD exhibit a *Lachnospiraceae*-associated gene**
342 **network involved in immune regulation**

343 In pairwise comparative analyses, patients with fibrostenotic CD (Montreal B2, $n=107$)
344 and patients using TNF- α -antagonists ($n=113$) exhibited several differentially abundant
345 microbial taxa. We therefore analyzed microbiota-associated host mucosal gene
346 interactions in these phenotypes (Figure 5). Pairwise comparisons between patients
347 with non-stricturing, non-penetrating disease vs. fibrostenotic CD revealed 2639
348 differentially abundant genes that were enriched in cellular energy metabolism and
349 immune system pathways (FDR<0.05, Supplementary Table S20). When comparing
350 microbial taxa, abundances of mucosal *Faecalibacterium*, *Erysipelotrichaceae_UCG-*
351 *003* and *Coprococcus_3* were lower in fibrostenotic CD, whereas abundances of
352 *Lachnospiraceae* and *Flavonifractor* were elevated in these patients (FDR<0.05). We
353 hypothesized that these altered bacterial abundances and gene expression patterns
354 may also translate into altered microbiota–gene networks relating to fibrostenotic CD. In
355 patients with non-stricturing, non-penetrating CD, we observed 1508 individual gene–
356 bacteria associations (corresponding to 84 different pathway–bacteria associations),
357 whereas we found 541 individual associations (corresponding to 40 different pathway–
358 bacteria associations) in patients with fibrostenotic CD. Comparing each bacteria–
359 associated gene cluster between patients with non-stricturing, non-penetrating and
360 fibrostenotic CD (FDR <0.05, Methods, Supplementary Table S21) identified four
361 distinct networks represented by mucosal *Lachnospiraceae*, *Coprococcus*,
362 *Erysipelotrichaceae* and *Flavonifractor*. The most significantly altered connections were
363 associated with *Lachnospiraceae*, which was associated with 955 genes in patients
364 with non-stricturing, non-penetrating CD, and these connections were mainly involved in

365 cell activation pathways such as vesicle-mediated cellular transport and membrane
366 trafficking (**Fig. 5A**). In total, 148 genes were associated with *Lachnoclostridium* in
367 patients with fibrostenotic CD (FDR<0.05), and these genes were involved in cellular
368 immunoregulatory interactions and adaptive immune system pathways (e.g. *CD8A*,
369 *CLEC2B* and *CXCR5*), tyrosine kinase signaling (e.g. *FGF16*), opioid signaling and G
370 alpha (s) signaling events (mediated via cAMP-dependent protein kinases, e.g. *POMC*,
371 *GNG7* and *GNG11*) and vesicle-mediated transport (e.g. *APOE*, *COLEC12* and *KIF3B*)
372 (**Fig. 5A-B**).

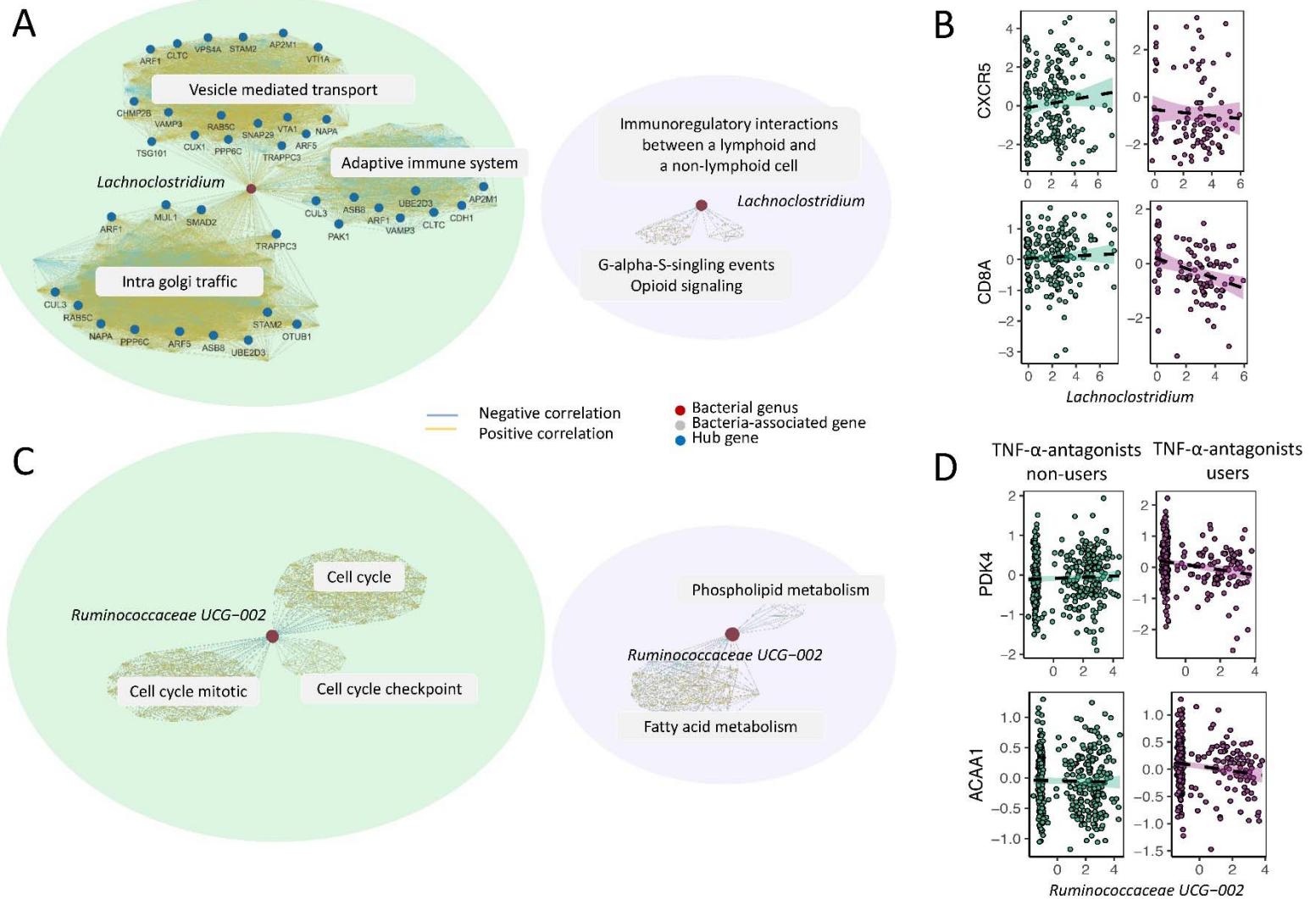
373 Earlier studies had shown that *Lachnoclostridium* bacteria are generally increased in
374 patients with (complicated) CD, e.g. postoperative CD [61], ASCA-positive CD [62] and
375 active granulomatous colitis [63]. Recently, *Lachnoclostridium* was also associated with
376 non-invasive diagnosis of colorectal adenoma and colorectal cancer [64,65]. These
377 associations may potentially explain associations with genes involved in cellular
378 proliferation and activation pathways. Increased abundances of *Lachnoclostridium* have
379 been observed in relation to pulmonary fibrosis and its progression [66] but not in
380 relation to intestinal fibrosis. In contrast, reduced abundances of *Faecalibacterium* and
381 *Eubacterium* species (belonging to the *Erysipelotrichaceae* family) have previously been
382 associated with luminal narrowing in patients with pediatric ileal CD [67]. Our results
383 suggest that it is not only increased *Lachnoclostridium* abundances that may play a role
384 in fibrostenotic CD, host immune-regulatory expression patterns may also vary along
385 with these bacterial shifts. Notably, as the tissues investigated in our study were not
386 derived from fibrotic regions, our findings show that these gene expression signatures
387 are already present in non-stenotic intestinal tissue.

388

389 **Use of TNF- α -antagonists is associated with *Ruminococcaceae*-associated gene
390 interactions related to fatty acid metabolism**

391 Subsequently, we investigated the impact of TNF- α -antagonist use on mucosal host–
392 microbe interactions. Pairwise comparisons revealed that TNF- α -antagonist use was
393 significantly associated with three different bacterial taxa, *Faecalibacterium*,
394 *Ruminococcaceae_UCG-002*, and *Ruminococcaceae_UCG-005* (all showing reduced

395 abundances in users), and 513 different genes (FDR<0.05, **Supplementary Table**
396 **S22**). For instance, one of the top genes associated with TNF- α -antagonist use was
397 CXCL13, which encodes B cell attracting chemokine 1. By comparing each tax-
398 associated gene cluster between patients using and not using TNF- α -antagonists, we
399 identified a single cluster represented by mucosal *Ruminococcaceae_UCG-002* that
400 was significantly altered in users vs. non-users (FDR<0.05, **Supplementary Table**
401 **S23**). *Ruminococcaceae_UCG-002* bacteria were associated with 135 genes in non-
402 users, and these genes were mainly enriched in cell cycle-associated pathways (e.g.
403 *PRIM1* and *PRIM2*), including mitosis-, prometaphase- and DNA-replication-associated
404 genes (**Fig. 5C-D**). However, the *Ruminococcaceae_UCG-002*-associated genes in
405 TNF- α -antagonist users (FDR<0.05) were predominantly involved in lipid/fatty acid
406 metabolism (e.g. *ACAA1*, *ACSL5* and *PDK4*), glycerophospholipid biosynthesis and
407 phospholipid metabolism. *Ruminococcaceae* comprise multiple distinct bacterial genera.
408 Some of these are part of the healthy gut microbiome [68], but others are potentially
409 pathogenic and commonly enriched in IBD [13,69]. The *Ruminococcaceae_UCG_002*
410 group is classified under the *Oscillospiraceae* family, which consists of obligate
411 anaerobic bacteria including *Faecalibacterium prausnitzii*. Depending on their micro-
412 environment, *Ruminococcaceae_UCG_002* bacteria can produce short-chain fatty acids
413 due to their fiber-metabolizing capacity [70-72]. The inverse associations between
414 *Ruminococcaceae_UCG_002* and genes involved in (peroxisomal) fatty acid oxidation
415 in patients using TNF- α -antagonists might reflect a beneficial therapeutic modulation,
416 i.e. a reduction of fatty acid oxidation and lipotoxicity, and possibly even attenuation of
417 microbiota-induced intestinal inflammation [73-85].



418

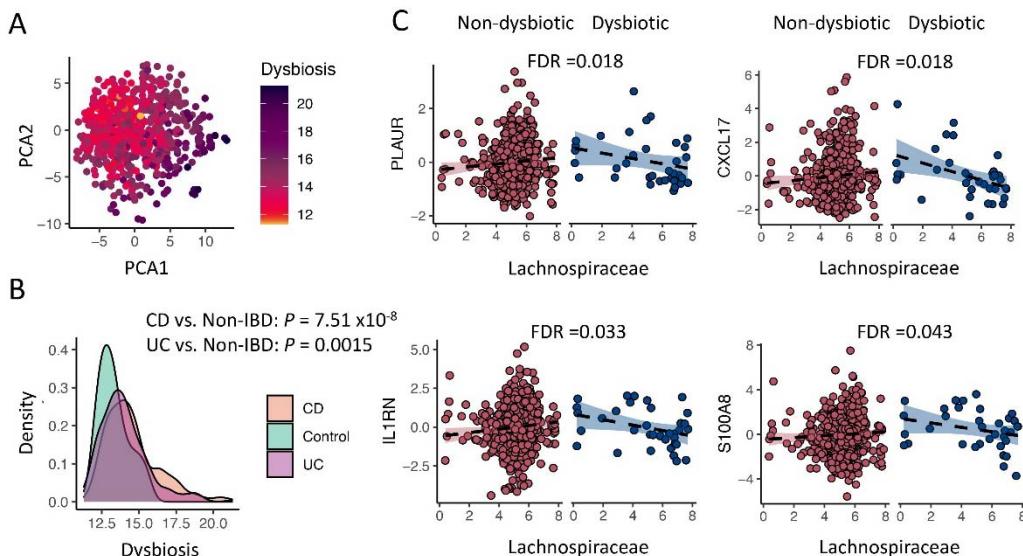
419 **Figure 5. Fibrostenotic CD and TNF- α -antagonist usage significantly alter mucosal host-microbe interactions in the context of IBD.**
 420 CentrLCC-network analyses were performed to characterize altered mucosal host-microbe interactions between different patient phenotypes.

421 Overall, fibrostenotic CD (Montreal B2 vs. non-stricturing, non-penetrating CD, i.e. Montreal B1) and use of TNF- α -antagonists (vs. non-users)
422 demonstrated significant modulation of observed mucosal host–microbe associations. **a**, Network graphs showing microbiota–gene association
423 networks in patients with non-stricturing, non-penetrating CD (Montreal B1) (left) and patients with fibrostenotic CD (Montreal B2) (right). When
424 comparing these patient groups, 5 bacterial taxa and 2639 host genes were significantly different (FDR<0.05). Four of the five bacterial taxa were
425 significantly altered in fibrostenotic CD vs. non-stricturing, non-penetrating CD, and *Lachnoclostridium* was the top bacteria involved (covering
426 63% of total associations in non-stricturing, non-penetrating CD and decreasing to 27% in fibrostenotic CD). In general, patients with fibrostenotic
427 CD were characterized by a loss of *Lachnoclostridium*–gene interactions. Red dots indicate gut microbiota. Blue dots indicate hub genes. Gray
428 fields indicate the main pathways represented by the associated genes. Yellow lines indicate positive associations between gene expression and
429 bacterial abundances. Light blue lines indicate negative associations. **b**, Key examples of *Lachnoclostridium*–gene interactions that were
430 significantly altered in patients with fibrostenotic CD compared to patients with non-stricturing, non-penetrating CD, including genes involved in
431 immunoregulatory interactions between lymphoid and non-lymphoid cells and tyrosine kinase signaling (*CD8A* and *CXCR5*). **c**, Network graphs
432 showing microbiota–gene interaction networks in patients not using TNF- α -antagonists (left) vs. patients using TNF- α -antagonists (right). When
433 comparing both groups, 3 bacterial groups and 513 genes were differentially abundant (FDR<0.05). Among these, a single bacterial group,
434 represented by *Ruminococcaceae_UCG_002*, was altered in interactions with host genes in patients using TNF- α -antagonists. **d**, Key examples of
435 *Ruminococcaceae_UCG_002*–gene interactions significantly altered in TNF- α -antagonists users vs. non-users. These genes were involved in
436 general biological processes such as the cell cycle but also included genes involved in fatty acid metabolism (*PDK4* and *ACAA1*).

437 **Mucosal host–microbe interactions depend on individual dysbiotic status**

438 As patients with IBD have microbial dysbiosis compared to healthy individuals, we
439 hypothesized that the strength and/or direction of the individual gene–bacteria
440 interactions may depend on the microbial community (eubiosis vs dysbiosis). We
441 therefore performed PCA on the microbiota data and calculated dysbiosis scores for all
442 patients and controls, as represented by the median Aitchison's distances to non-IBD
443 controls (**Fig. 6A**). Patients with IBD demonstrated higher dysbiosis scores compared to
444 controls (CD vs. non-IBD: $P=5.1\times10^{-8}$, UC vs. non-IBD: $P=0.0015$), but there was no
445 clear difference between patients with CD and UC (**Fig. 6B**). When comparing patients
446 with IBD above and below the 90th percentile of dysbiosis scores [13], 204 individual
447 gene–bacteria interactions showed significant dependence on microbial dysbiosis (**Fig.**
448 **6C, Supplementary Table S24**) ($FDR<0.05$). We also performed permutation tests,
449 which confirmed that the significant interactions were not observed by chance
450 (**Methods**, $FDR<0.05$). In one example of these interactions, expression of the *PLAUR*
451 gene encoding for the urokinase plasminogen activator surface receptor was positively
452 associated with *Lachnospiraceae* abundance, but this shifted to an inverse association
453 when only considering individuals with a high degree of mucosal dysbiosis (90–100%)
454 ($P=1.69\times10^{-6}$). The Ly6/PLAUR domain containing protein 8 (Lypd8) also functions as
455 an antimicrobial peptide and has previously been shown to be capable of protecting the
456 host from invading pathogenic flagellated bacteria [86]. Another example is the positive
457 association between *S100A8*, which encodes S100 calcium-binding protein A8 (also
458 known as calgranulin A), and *Lachnospiraceae*, which showed a negative association in
459 individuals with high dysbiosis ($P=1.78\times10^{-5}$). *S100A8* has a wide variety of functions in
460 regulating inflammatory processes and forms a heterodimer with *S100A9*, also known
461 as calprotectin, which is used as a biomarker for inflammatory activity in IBD. Its known
462 antimicrobial activity towards bacteria via chelation of zinc ions, which are essential for
463 microbial growth, may therefore be disrupted in a dysbiotic environment [44]. Similar to
464 the two previous examples, the observed association between the expression of *IL1RN*
465 (encoding for the interleukin-1 receptor antagonist protein) and *Lachnospiraceae* shifted
466 from positive to negative ($P=4.10\times10^{-5}$), indicating that the natural protection against the
467 proinflammatory effects of IL-1 β , which associates with *Lachnospiraceae* abundance,

468 may be lost in circumstances of high microbial dysbiosis. Similarly, expression of the
469 CXCL17 gene encoding for a mucosal chemokine protein known to exert broad
470 antimicrobial activity [87] positively correlated with *Lachnospiraceae* abundance, which
471 was clearly different among individuals with higher dysbiosis scores.



472

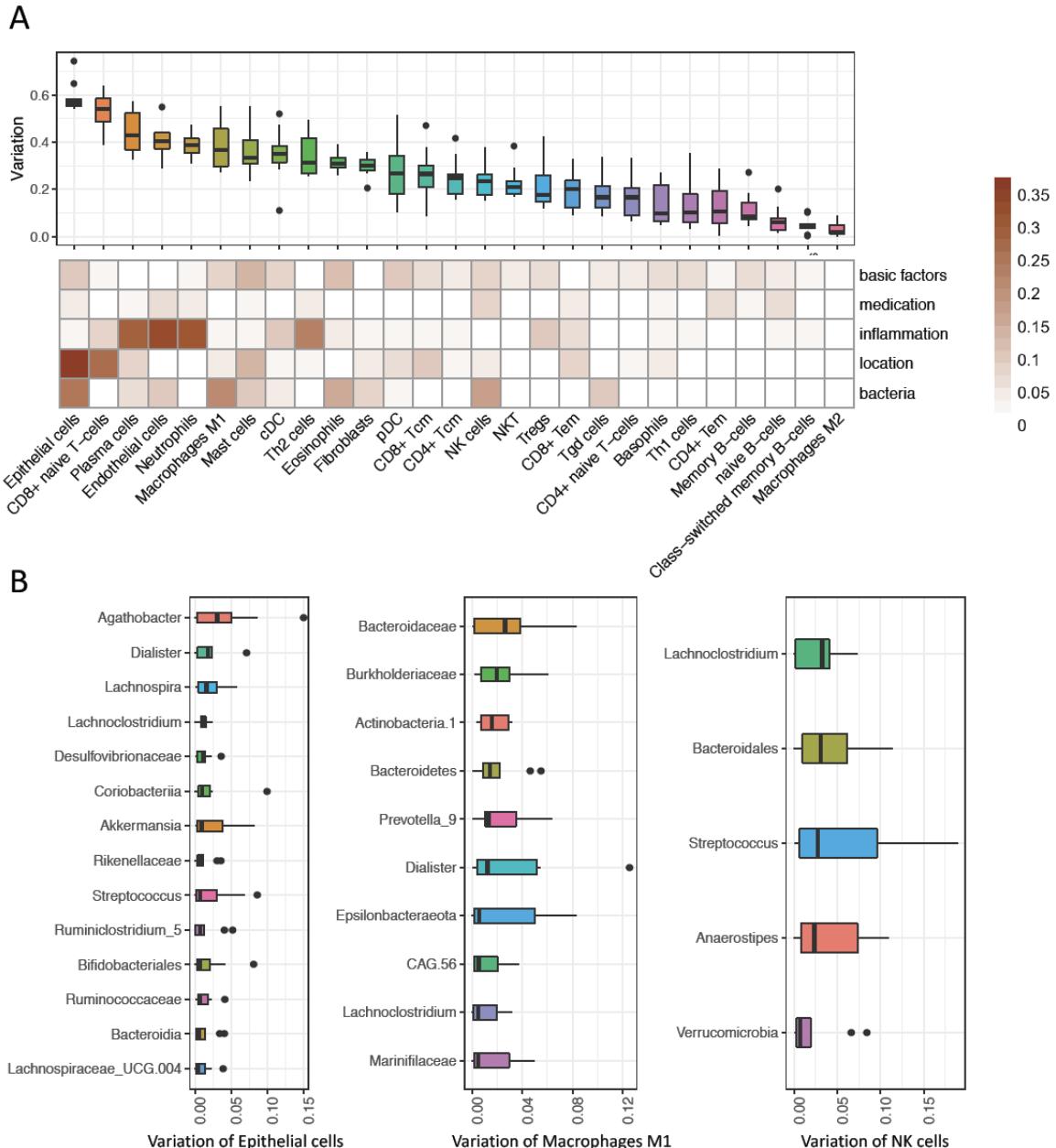
473 **Figure 6. Mucosal host-microbe interactions depend on individual dysbiotic status.** **a**, PCA of
474 mucosal 16S rRNA sequencing data shows that degree of mucosal dysbiosis explains a large part of
475 microbial variation. **b**, Dysbiosis scores were generally higher among patients with CD and UC compared
476 to controls. **c**, Key examples of individual gene–bacteria interactions that demonstrate a directional shift
477 upon higher dysbiosis (90–100%) as compared to patients with lower dysbiosis scores (0–90%). Mucosal
478 *Lachnospiraceae* bacteria positively associate with the expression of the *PLAUR*, *CXCL17*, *IL1RN* and
479 *S100A8* genes, whose gene products all have beneficial antimicrobial activity towards pathogenic
480 bacteria. CD, Crohn’s disease. PCA, principal component analysis. UC, ulcerative colitis.

481

482 **Mucosal microbiota associate with variation in intestinal cell type–enrichment**

483 Subsequently, we aimed to evaluate which intestinal cell types are involved in mucosal
484 host–microbe interactions (**Figure 7, Extended Data Fig. S7**). Deconvolution of host
485 gene expression data revealed that the mucosal microbiota was significantly associated
486 with several cell types, but most evidently with intestinal epithelial cells, M1
487 macrophages, NK cells and mucosal eosinophils. Tissue inflammatory status and

488 location also strongly contributed to the variation in most intestinal cell types. Mucosal
489 microbiota that were significantly associated with intestinal epithelial cell enrichment
490 typically belonged to the Firmicutes phylum, including *Agathobacter*, *Dialister*,
491 *Lachnospira*, *Lachnoclostridium* and *Ruminococcaceae* (**Supplementary Table S25**).



492
493 **Figure 7. Mucosal microbiota associate with distinct intestinal mucosal cell types.** a, Boxplots show
494 the amount of variation in intestinal cell type-enrichment that could be explained by mucosal microbiota.
495 Heatmap shows the contribution of other fitted models in explaining intestinal cell type-enrichment,

496 including 'basic factors' (age, sex and BMI), medication use, tissue inflammatory status and tissue
497 location. Mucosal microbiota contributed most to the variation in enrichment of intestinal epithelial cells,
498 M1-macrophages, NK cells and eosinophils. **b**, Boxplots showing the contribution of the main bacterial
499 taxa that explain the variation in mucosal enrichment of intestinal epithelial cells, M1-macrophages and
500 NK cells—the cell types that interacted most strongly with the mucosal microbiota.

501 **Discussion**

502 In this study, we show distinct mucosal host–microbe interactions in intestinal tissue
503 from patients with IBD. Mucosal gene expression patterns in IBD are mainly determined
504 by tissue location and inflammatory status and systematically demonstrate upregulation
505 of distinct inflammation-associated genes, even in endoscopically non-inflamed tissue.
506 Subsequently, we observed that the mucosal microbiota composition in patients is
507 marked by high inter-individual variability. The main focus of our analyses, however,
508 was integrative analysis of both data entities, which allowed us to comprehensively
509 uncover many host–microbe associations, both on component level and as individual
510 associations in IBD. Furthermore, we identify specific transcriptional networks that are
511 significantly altered in patients with fibrostenotic CD and patients using TNF- α -
512 antagonists and observe that these associations depend on the degree of mucosal
513 dysbiosis. Finally, we show that mucosal microbiota are significantly associated with
514 intestinal cell type composition, in particular with epithelial cells, macrophages and NK
515 cells.

516 Tissue location and inflammatory status have the greatest impact on the variation in
517 mucosal gene expression patterns. Enriched genes are mainly represented by those
518 involved in pathophysiological pathways relevant to IBD, e.g. interleukin and interferon
519 signaling and ECM remodeling. Patients with CD and UC show striking differences, e.g.
520 Notch-1 signaling pathways are upregulated in CD, while genes involved in nutrient
521 absorption and lipid metabolism are downregulated. Activation of Notch-1 signaling has
522 been associated with improved mucosal barrier function, driven by lamina propria-
523 residing CD4 $^{+}$ -T-lymphocytes that induce intestinal epithelial cell differentiation [17].
524 Notch-1 signaling more efficiently spreads within CD intestinal epithelia, as compared to
525 UC or control epithelia. Notch-1 is not only implicated in IBD, it also confers protection
526 against the development of colorectal carcinoma via p53 signaling, thereby promoting
527 cell cycle arrests and cellular apoptosis [18,88,89]. Since UC patients with long-lasting
528 colonic inflammation have a higher risk of developing IBD-associated colorectal
529 carcinoma, we hypothesize that downregulation of Notch-1 in these patients may
530 potentially be involved in carcinogenesis.

531 Analysis of mucosal microbiota in patients with IBD reveals reduced alpha-diversity,
532 microbial dissimilarity and marked intra-individual variability that is particularly strong in
533 CD but still present to a lesser extent in UC. Given the large heterogeneity in IBD and
534 the fact that compositional differences are largely attributable to individual phenotypic
535 factors, cautious interpretation is warranted when associating mucosal microbial profiles
536 to disease phenotypes or outcomes, rendering them inappropriate for diagnostic
537 purposes. These observations corroborate those of previously published mucosal 16S
538 studies in IBD [13,21,22]. Moreover, our findings align with a recent prospective meta-
539 analysis study that concluded there is sparse evidence for additional population
540 structure in mucosal microbiomes in IBD, e.g. microbiota-driven discrete disease
541 subtypes within IBD [90].

542 Sparse-CCA analysis was performed to capture the key pathway–bacteria interactions.
543 These include numerous inverse associations between bifidobacteria and expression of
544 genes involved in fatty acid metabolism, which align well with previously published data
545 from animal studies demonstrating anti-inflammatory and anti-lipogenic effects of
546 *Bifidobacterium* treatment on chemically-induced intestinal inflammation [29,31,32]. For
547 example, treatment with *Bifidobacterium adolescentis* IM38 attenuated high fat diet–
548 induced colitis in mice by inhibiting lipopolysaccharide production, NF-κB activation and
549 TNF-expression in colonic epithelial cells [31]. Likewise, treatment with *Bifidobacterium*
550 *infantis*, with or without a combination of inulin-type fructans, ameliorated DSS-induced
551 colitis in rats, as evidenced by decreased expression of IL-1 β , malondialdehyde (MDA,
552 a lipid peroxidation marker), decreased bacterial translocation and increased production
553 of short-chain fatty acids [32]. In line with our findings, this supports the ongoing quest
554 for efficacious probiotic (bifidobacteria-containing) interventions in patients with IBD
555 [91,92]. In addition, we observe a *Ruminococcaceae*-UCG-002-associated network of
556 genes involved in (peroxisomal) fatty acid oxidation and lipotoxicity, which are inversely
557 associated with these bacteria in patients using TNF- α -antagonists. Interestingly,
558 multiple studies have observed that *Ruminococcaceae* increase after anti-TNF therapy
559 in patients with CD and UC [73,75-77]. One of these studies specifically identified an
560 association between the *Ruminococcaceae*_UCG-002 group and responsiveness to
561 TNF- α -antagonists, albeit not in relation to host gene expression patterns [75].

562 Strikingly, many of the network-associated genes we observe are controlled by the
563 PPAR- γ transcription factor, a butyrate sensor that may result in reduced lipotoxicity and
564 reduced intestinal inflammation through prevention of overgrowth of potentially
565 pathogenic bacteria [79-85]. These findings underscore the potential relevance of
566 PPAR- γ as a therapeutic target in IBD [85].

567 We also observed an intriguing inverse relationship between *Erysipelotrichaceae* and
568 intestinal ECM remodeling pathways, which may support the notion that intestinal
569 fibrosis in IBD is highly linked to microbial composition [60,93,94]. Interestingly, a
570 decreased relative abundance of *Erysipelotrichaceae* has previously been observed in
571 patients with collagenous colitis [55] and cystic fibrosis-related lung fibrosis [56-58], as
572 well as in mice with liver fibrosis and hepatocyte-specific *NOD2* deletions [59]. In CD,
573 several bacterial species belonging to *Erysipelotrichaceae*, including *Clostridium*
574 *innocuum* and *Erysipelotrichaceae ramosum*, have been associated with the expansion
575 of mesenteric adipose tissue (“creeping fat”), a unique feature of CD [60]. Creeping fat
576 in CD has previously been characterized by higher abundances of *Erysipelotrichaceae*
577 compared to adjacent mesenteric adipose tissue and underlying mucosal tissue and is
578 accompanied by higher expression of ECM- and collagen-related genes. *C. innocuum*
579 translocated to mesenteric fat, promoted fibrosis and stimulated tissue-remodeling in
580 patients with CD, resulting in an adipose tissue barrier that may prevent systemic
581 translocation of intestinal bacteria [60]. This phenomenon could potentially explain the
582 inverse associations we observe between expression of ECM remodeling and mucosal
583 *Erysipelotrichaceae*. In our differential network analyses, we observe a substantial
584 decrease of *Lachnoclostridium*-associated genes in patients with fibrostenotic CD that
585 are mainly associated with cellular immunoregulatory interactions and adaptive immune
586 system pathways. These findings suggest that *Lachnoclostridium*-associated
587 immunoregulatory expression patterns may play a role in fibrostenotic CD. Although
588 little is known about the exact role of *Lachnoclostridium* in IBD, these bacteria were
589 recently strongly associated with the development of colorectal cancer and with
590 pulmonary fibrosis [64-66].

591 Another key host–microbe interaction module pertains to *Bacteroides*, which inversely
592 correlates with interleukin signaling and positively associates with metal stress response
593 transcription factors encoding for MTs. To maintain cellular redox balance, MTs detoxify
594 heavy metal ions and scavenge ROS, thereby attenuating oxidative stress. Previous
595 studies have shown that MTs may prevent experimental colitis or act as danger signals
596 by mediating immune cell infiltration in the intestine [45,46]. Although experimental
597 evidence seems to be inconclusive, there is ample evidence indicating a role for
598 aberrant MT homeostasis in IBD [47]. This mechanism depends on the intracellular
599 accumulation of zinc, which induces autophagy under chronic NOD2-stimulation. In IBD,
600 the mucosal microbiota may contribute to the regulation of MT expression, intracellular
601 zinc homeostasis and autophagy, thereby regulating intracellular bacterial clearance by
602 intestinal macrophages. Findings from this study may support a putative role for
603 *Bacteroides* in modulating MT activation, thereby contributing to intracellular redox
604 homeostasis, zinc levels, macrophage autophagy, or even host defense against
605 pathogens. Importantly, MTs and zinc regulation constitute potential therapeutic targets
606 in IBD [44-47, 95-97].

607 Individual gene–bacteria association analysis revealed distinct mucosal host–microbe
608 interactions that largely overlap with those from the sparse-CCA analysis, but these
609 provide more granular insight into the observed associations. Key examples of
610 individual host gene–bacteria interactions are listed in **Box 1**. Amongst others, we
611 demonstrate several host–microbe interactions that are putatively involved in
612 immunological tolerance and prevention of autoimmunity (e.g. bifidobacteria and
613 *FOSL1/KLF2* expression), colorectal carcinogenesis (e.g. *Anaerostipes* and *SMAD4*,
614 *Akkermansia* and *YDJC*) and inflammatory signaling (e.g. *Oscillibacter* and *OSM*
615 expression). Notably, many of these associations are dependent on fibrostenotic
616 disease, TNF- α -antagonist use and the degree of mucosal dysbiosis. In addition,
617 deconvolution of the mucosal RNA-seq data reveal cell type–specific patterns of
618 microbial interactions that warrant further study, for example through single-cell RNA-
619 seq studies.

620 Mucosal host–microbiota interactions have been investigated previously in both cohort
621 (e.g. the HMP2 and Irish IBD) and experimental studies [12-16]. Alongside several
622 observations consistent with previous findings, we identify many novel host–microbe
623 interactions. Differences in sample size, patient phenotypes and sample handling may
624 be at least partially responsible for these observations. In our study, large groups of
625 gene–bacteria associations are revealed that cover a wide range of molecular
626 mechanisms potentially relevant in the context of IBD, including immune response
627 pathways, cellular processes and a variety of metabolic pathways. Moreover, our study
628 features the largest sample size so far [12-15], and this enabled us to perform an
629 integrative analysis with respect to the large disease heterogeneity and identify novel
630 host–microbiota crosstalk related to different clinical characteristics. However, several
631 limitations also warrant recognition. As our study is of cross-sectional origin, we cannot
632 assess the longitudinal dynamics of host–microbe interactions to discover signatures for
633 therapy responsiveness or disease prognosis. Consequently, our associative results
634 cannot establish potential causality between microbial abundances and host gene
635 expression. Functional experiments are thus required to validate the biological
636 relevance of the individual host–microbe interactions, as well as their behavior in
637 microbial ecosystems. Finally, bowel preparation prior to the endoscopic procedure or
638 cross-contamination between biopsy sites during endoscopy can affect the mucosal
639 microbiota composition [21,50,98].

640 Our results demonstrate a complex and heterogeneous interplay between mucosal
641 microbiota and mucosal gene expression patterns that is concomitant with the strong
642 impact of specific patient traits in a large cohort of patients with IBD. Our findings may
643 guide development of mechanistic studies (e.g. host–microbe co-culture systems) that
644 could provide functional confirmation of relevant pathophysiological gene–bacteria
645 interactions and serve as a resource for rational selection of therapeutic targets in IBD.
646 This study presents a large-scale, comprehensive landscape of intestinal host–microbe
647 interactions in IBD that could aid in guiding drug development and provide a rationale
648 for microbiota-targeted therapy as a strategy to control disease course. Future studies
649 are warranted to focus on the integration of host–microbe interaction modules in
650 prospective clinical trials investigating their utility for predicting disease course and

651 responsiveness to treatment and for stratifying patients to facilitate therapeutic decision-
652 making.

653 **Methods**

654

655 **Study population**

656 Patients with an established diagnosis of IBD were included at the outpatient clinic of
657 the University Medical Center Groningen (UMCG) based on their participation in the
658 1000IBD project, for which detailed phenotypic information and multi-omics profiles had
659 been generated [99]. Patients included in this study were at least 18 years old and were
660 enrolled from 2003–2019. Diagnosis of IBD was based upon clinical, laboratory,
661 endoscopic and histopathological criteria, with the latter criteria also used to determine
662 the inflammatory status of collected biopsies. Detailed phenotypic data were collected
663 for all patients, including age, sex, BMI (body weight divided by squared height),
664 smoking status, Montreal disease classification, medication usage, history of surgery,
665 clinical disease activity and histological disease activity, and all were assessed at time
666 of sampling. Montreal disease classification was recorded from the closest visit to the
667 outpatient clinic at time of sampling. Clinical disease activity was established using the
668 Harvey-Bradshaw Index (HBI) for patients with CD and the Simple Clinical Colitis
669 Activity Index (SCCAI) for patients with UC. We further included 17 healthy non-IBD
670 controls ($n=59$ biopsies) who underwent endoscopy because of clinical suspicion of
671 intestinal disease or within the context of colon cancer screening. All participants
672 provided written informed consent prior to sample collection. This study was approved
673 by the Institutional Review Board (IRB) of the UMCG, Groningen, the Netherlands (in
674 Dutch: ‘Medisch Ethische Toetsingscommissie’, METc; IRB nos. 2008/338 and
675 2016/424) and was conducted according to the principles of the Declaration of Helsinki
676 (2013).

677

678 **Mucosal RNA-sequencing**

679 711 intestinal biopsies from 420 patients with IBD were collected. These were
680 immediately snap-frozen in liquid nitrogen by an endoscopy nurse or research
681 technician present during the endoscopic procedure. Biopsy inflammatory status was

682 assessed based on histological examination by certified pathologists. Biopsies were
683 stored at -80°C until further processing.

684 RNA isolation was performed using the AllPrep DNA/RNA mini kit (Qiagen, reference
685 number: 80204) according to manufacturer's instructions. Homogenization of intestinal
686 biopsies was performed in RLT lysis buffer including β -mercaptoethanol using the
687 Qiagen Tissue Lyser with stainless steel beads (diameter 5 mm, reference number:
688 69989). For the first sample batch, sample preparation was executed using the
689 BioScientific NEXTflexTM Rapid Directional RNA-Seq Kit (Perkin-Elmer). Paired-end
690 sequencing of RNA was performed using the Illumina NextSeq500 sequencer (Illumina).
691 For the second sample batch, sample preparation was performed for construction of the
692 Eukaryotic Transcriptome Library (Novogene). Paired-end sequencing of RNA was
693 performed using the Illumina HiSeq PE250 platform. Sequencing was performed in two
694 different batches, which necessitated pseudo-randomization (covering type of IBD
695 diagnosis, biopsy location and disease activity) across plates to mitigate potential batch
696 effects. The batch effects have been taken into account in all the analysis. On average,
697 approximately 25 million reads were generated per sample.

698 Raw read quality was checked using FastQC with default parameters (ref v.0.11.7).
699 Adaptors identified by FastQC were clipped using Cutadapt (ref v1.1) with default
700 settings. Sickle (ref v1.200) was used to trim low-quality ends from the reads (length
701 <25 nucleotides, quality <20). Reads were aligned to the human genome
702 (human_glk_v37) using HISAT (ref v0.1.6) (with maximum allowance of two
703 mismatches), and read sorting was performed using SAMtools (ref v0.1.19). SAMtools
704 flagstat and Picard tools (ref v2.9.0) were used to obtain mapping statistics. Six samples
705 with low percentage read alignment (< 90%) were removed. Gene expression was
706 estimated using HTSeq (ref v0.9.1), based on Ensemble version 75 annotation,
707 resulting in a RNA expression dataset featuring 15,934 genes. Expression data on gene
708 level were normalized using a trimmed mean of M values, and c/r transformation was
709 applied, resulting in 826 mucosal RNA-seq samples.

710

711 **Mucosal 16S rRNA gene sequencing**

712 Total DNA extraction of intestinal biopsies using 0.25 g of sample was performed as
713 described previously, with minor modifications [100]. Microbial composition of intestinal
714 biopsies was determined by Illumina MiSeq paired-end sequencing of the V3-V4
715 hypervariable region of the 16S rRNA gene (MiSeq Benchtop Sequencer, Illumina Inc.,
716 San Diego, USA). Amplification of bacterial DNA was performed by PCR using modified
717 341F and 806R primers with a six-nucleotide barcode on the 806R primer for
718 multiplexing [101,102]. Sequences of both primers can be found in **Supplementary**
719 **Table S1**. Both primers contain an Illumina MiSeq adapter sequence, which is
720 necessary for flow cell-binding in the MiSeq machine. A detailed overview of the PCR,
721 DNA clean-up and MiSeq library preparation using a 2x300 cartridge can be found in
722 the **Supplementary Methods**. Read trimming and filtering was done using
723 *Trimmomatic* (0.33) to obtain an average read quality of 25 and a minimum length of 50.
724 Quality was further checked using R package *DADA2* (v1.03) with the following
725 parameters: truncLen=c(240,160), maxN=0, maxEE=c(2,2), truncQ=2 and
726 rm.phix=TRUE. After error correction and chimera removal, the amplicon sequence
727 variants were assigned to the *silva* database (v.132). Samples with >2,000 mapped
728 reads were used for further analysis, resulting in 755 mucosal 16S samples. After
729 accounting for overlap between mucosal RNA-seq and mucosal 16S data, 696 intestinal
730 biopsies from 337 different patients and 16 non-IBD controls were available for host–
731 microbiota interaction analyses.

732

733 **Statistical analysis**

734 *Descriptive statistics*

735 Descriptive data are presented as means \pm standard deviation (SD), medians
736 [interquartile range, IQR] or proportions n with corresponding percentages (%).
737 Between-group comparisons were performed using Mann-Whitney *U*-tests, Pearson's
738 chi-squared tests or Fisher's exact tests (if n observations were <10). Nominal *P*-values
739 ≤ 0.05 were considered statistically significant.

740 *Mucosal gene expression analysis*

741 Sample gene expression dissimilarity was calculated using Aitchison's distances.
742 General linear models were used to assess the associations between mucosal gene
743 expression and clinical phenotypes while controlling for potential confounders, which
744 were determined from our previous study (medication included the use of
745 aminosalicylates, thiopurines and steroids) [103]. In particular, to assess the effect of
746 mucosal inflammation on gene expression, we re-coded the inflammation status in an
747 ordinal fashion as 0, 1 or 2 to represent biopsies from non-IBD controls, biopsies from
748 non-inflamed tissue of patients with IBD and biopsies from inflamed areas of patients
749 with IBD, respectively. Intestinal inflammatory status was thus treated as a continuous
750 variable to account for presence of residual inflammation in biopsies marked as being
751 taken from non-inflamed areas in the intestines. A correction for multiple hypotheses
752 testing was applied using an FDR threshold of 5%.

753 1) Inflammation-associated genes were identified in three comparisons: (1)
754 CD colonic inflamed tissue vs. CD colonic non-inflamed tissue vs. non-IBD
755 colonic tissue, (2) CD ileocecal inflamed tissue vs. CD ileocecal non-inflamed
756 tissue vs. non-IBD ileocecal tissue and (3) UC colonic inflamed tissue vs. UC
757 colonic non-inflamed tissue vs. non-IBD colonic tissue:

758 *Gene ~ intercept + inflammation + age + sex + BMI + medication + batch*

759 2) Clinical phenotype-associated genes were identified using the following
760 model:

761 *Gene ~ intercept + Montreal/anti-TNF therapy + age + sex + BMI + inflammation*
762 *+ tissue location + medication + batch*

763

764 *Microbial characterization*

765 Microbial richness and evenness was determined by calculating the Shannon index
766 representing alpha-diversity of the gut microbiota. Microbial dissimilarity of samples was
767 determined by calculating Aitchison's distances after *clr* transformation using the R
768 package *Compositions* (v2.02). Analysis of paired samples from the same individuals
769 was performed while comparing microbial features between inflammation status,

770 disease location and disease subtype using paired Wilcoxon tests. Factors potentially
771 influencing mucosal microbiota were determined using Hierarchical All-against-All
772 significance testing (HALIA) [104]. Associations between microbial features and biopsy
773 inflammatory status, IBD diagnosis, disease location (biopsy origin) and clinical
774 phenotypes were performed using general linear models (see below). Per sample, the
775 mucosal dysbiosis score was defined as the median Aitchison distance from that
776 sample to a reference sample set of non-IBD controls. Dysbiotic status was defined as
777 being at the 90th percentile of this score [13].

778 1) Associations between microbial taxa and biopsy inflammation/location:

779 *Taxa ~ intercept + inflammation + location + age + sex + BMI + medication +*
780 *batch + surgical resection*

781 2) Associations between microbial taxa and clinical phenotypes:

782 *Taxa ~ intercept + Montreal/anti-TNF therapy + inflammation + location + age +*
783 *sex + BMI + medication + batch + surgical resection*

784

785 *Gene–microbiota interaction analysis*

786 We first focused on host inflammation-related genes ($n=1,437$) to investigate their
787 potential associations with mucosal microbiota. Group-level correlations between gene
788 expression and mucosal microbiota were performed using sparse-CCA using the
789 residuals of genes and microbiota after correcting for age, gender, BMI, inflammation,
790 tissue location and surgical resection separately. Sparse-CCA identifies the PCs from
791 two related datasets that maximize the correlation between the two components. A set
792 of enriched host pathways for all significant components was combined while adjusting
793 for multiple comparisons using the FDR approach. Individual pairwise gene–microbiota
794 associations were assessed by fitting a general linear model while adjusting for age,
795 sex, BMI, inflammation status, tissue location, sequencing batch and medication use
796 (including the use of aminosalicylates, thiopurines and steroids, see below). A gene–
797 microbiota network analysis was visualized using the R package *ggview*.

798 1) Individual gene–bacteria associations were determined using the following
799 model:

800 *Gene ~ intercept + taxa + inflammation + location + age + sex + BMI +*
801 *medication + batch*

802 Second, we focused on host–microbiota interactions associated with fibrostenotic CD
803 and usage of TNF- α -antagonists. Genes and taxa that were differentially abundant
804 between clinical phenotypes were selected and then served as input for CentrLCC-
805 network analysis using the *NetCoMi* R package. Hub nodes were defined as those with
806 an eigenvector centrality value above the empirical 95% quantile of all eigenvector
807 centralities in the network. This analysis was done in different groups separately (e.g.
808 users and non-users of TNF- α -antagonists). To assess whether the taxa-associated
809 gene networks were altered between groups, the associated genes for each taxa node
810 were ranked within the total geneset background based on Z-scores. The Wilcoxon test
811 was used to compare the two gene rank lists for each taxa.

812 Third, we assessed whether gene–microbiota associations depend on intestinal
813 dysbiosis by modeling these associations using an additional interaction term in linear
814 models. The dysbiosis score was treated as a continuous value. To determine whether
815 these interactions were observed by chance, we also performed permutation tests that
816 randomly shuffled the dysbiosis score 100 times across all samples, and then repeated
817 the interaction models. On average, only three FDR-adjusted significant results were
818 obtained for each round of permutation testing, suggesting that the rate of total false
819 positives was approximately ~ 0.014 (3/204).

820 2) *Gene ~ intercept + taxa + dysbiosis + taxa * dysbiosis + inflammation +*
821 *location + age + sex + BMI + medication + batch*

822 Fourth, enrichment of specific intestinal cell types was inferred from the RNA-seq data
823 using the *Xcell* package in R. The effects of tissue location, inflammatory status and
824 type of IBD diagnosis on expression levels of mucosal cell types were assessed using
825 linear models, adjusting for age, sex, BMI, batch and medication usage. Subsequently,
826 we used the *glmnet* R package to investigate the variation of cell type–enrichment that

827 could be explained by the mucosal microbiota using *lasso* regression while employing a
828 nested 10-fold cross-validation using six models:

829 1) *Cell enrichment* ~ age + gender + BMI + batch

830 2) *Cell enrichment* ~ medication (aminosalicylates, thiopurines, steroids,
831 biologicals)

832 3) *Cell enrichment* ~ inflammation

833 4) *Cell enrichment* ~ tissue location

834 5) *Cell enrichment* ~ bacteria abundance

835 6) *Cell enrichment* ~ full factors mentioned above

836 The percentage of explained variance (R^2) was calculated to estimate the variation in
837 cell type–enrichment explained by the mucosal microbiota. All analyses were corrected
838 for multiple testing using a FDR significance threshold of 0.05. All gene pathway
839 enrichment analyses were conducted using the Reactome database from MsigDB
840 [105,106].

841

842 *Replication in the HMP2 dataset*

843 RNA-seq and 16S raw data were obtained from <https://ibdmdb.org> and reprocessed
844 using the same pipeline in this study. After harmonizing with the phenotype file, we
845 included 152 intestinal biopsies from the 85 patients with CD, 46 patients with UC and
846 45 non-IBD controls. First, gene expression and mucosal microbiota patterns were
847 compared separately between this study and HMP2. Second, given the limited overlap
848 in clinical phenotypes between the two cohorts, we restricted the replication analysis to
849 inflammation-related host–microbiota interactions. Individual gene–microbiota
850 associations were calculated using the same linear models used in this study while
851 adjusting for age, gender, tissue location and inflammation status. Spearman correlation
852 coefficients were used to assess the concordance between the Z-scores of gene–
853 microbiota associations from the two studies.

854 **References**

855 1. Chang JT. Pathophysiology of Inflammatory Bowel Diseases. *N Engl J Med*
856 2020;383(27):2652-2664.

857 2. de Lange KM, Moutsianas L, Lee JC, Lamb CA, Luo Y, Kennedy NA, et al. Genome-
858 wide association study implicates immune activation of multiple integrin genes in
859 inflammatory bowel disease. *Nat Genet* 2017;49(2):256-261.

860 3. de Souza HSP, Fiocchi C, Iliopoulos D. The IBD interactome: an integrated view of
861 aetiology, pathogenesis and therapy. *Nat Rev Gastroenterol Hepatol* 2017;14(12):739-
862 749.

863 4. Franzosa EA, Sirota-Madi A, Avila-Pacheco J, Fornelos N, Haiser HJ, Reinker S, et
864 al. Gut microbiome structure and metabolic activity in inflammatory bowel disease. *Nat*
865 *Microbiol* 2019;4(2):293-305.

866 5. Vich Vila A, Imhann F, Collij V, Jankipersadsing SA, Gurry T, Mujagic Z, et al. Gut
867 microbiota composition and functional changes in inflammatory bowel disease and
868 irritable bowel syndrome. *Sci Transl Med* 2018;10(472):eaap8914.

869 6. Frank DN, St Amand AL, Feldman RA, Boedeker EC, Harpaz N, Pace NR.
870 Molecular-phylogenetic characterization of microbial community imbalances in human
871 inflammatory bowel diseases. *Proc Natl Acad Sci U S A* 2007;104(34):13780-5.

872 7. Kostic AD, Xavier RJ, Gevers D. The microbiome in inflammatory bowel disease:
873 current status and the future ahead. *Gastroenterology* 2014;146(6):1489-99.

874 8. Kurilshikov A, Wijmenga C, Fu J, Zhernakova A. Host Genetics and Gut Microbiome:
875 Challenges and Perspectives. *Trends Immunol* 2017;38(9):633-647.

876 9. Hu S, Vich Vila A, Gacesa R, Collij V, Stevens C, Fu JM, et al. Whole exome
877 sequencing analyses reveal gene-microbiota interactions in the context of IBD. *Gut*
878 2021;70(2):285-296.

879 10. Cohen LJ, Cho JH, Gevers D, Chu H. Genetic Factors and the Intestinal Microbiome
880 Guide Development of Microbe-Based Therapies for Inflammatory Bowel Diseases.
881 *Gastroenterology* 2019;156(8):2174-2189.

882 11. Davison JM, Lickwar CR, Song L, Breton G, Crawford GE, Rawls JF. Microbiota
883 regulate intestinal epithelial gene expression by suppressing the transcription factor
884 Hepatocyte nuclear factor 4 alpha. *Genome Res* 2017;27(7):1195-1206.

885 12. Priya S, Burns MB, Ward T, Mars RAT, Adamowicz B, Lock EF, et al. Identification
886 of shared and disease-specific host gene-microbiome interactions across human
887 diseases using multi-omic integration. *Nat Microbiol* 2021; doi: 10.1038/s41564-022-
888 01121-z.

889 13. Lloyd-Price J, Arze C, Ananthakrishnan AN, Schirmer M, Avila-Pacheco J, Poon
890 TW, et al. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases.
891 *Nature* 2019;569(7758):641-648.

892 14. Häslер R, Sheibani-Tezerji R, Sinha A, Barann M, Rehman A, Esser D, et al.
893 Uncoupling of mucosal gene regulation, mRNA splicing and adherent microbiota
894 signatures in inflammatory bowel disease. *Gut* 2017;66(12):2087-2097.

895 15. Morgan XC, Kabakchiev B, Waldron L, Tyler AD, Tickle TL, Milgrom R, et al.
896 Associations between host gene expression, the mucosal microbiome, and clinical
897 outcome in the pelvic pouch of patients with inflammatory bowel disease. *Genome Biol*
898 2015;16(1):67.

899 16. Lipinski S, Till A, Sina C, Arlt A, Grasberger H, Schreiber S, et al. DUOX2-derived
900 reactive oxygen species are effectors of NOD2-mediated antibacterial responses. *J Cell*
901 *Sci* 2009;122(Pt 19):3522-30.

902 17. Dahan S, Rabinowitz KM, Martin AP, Berin MC, Unkeless JC, Mayer L. Notch-1
903 signaling regulates intestinal epithelial barrier function, through interaction with CD4+ T
904 cells, in mice and humans. *Gastroenterology* 2011;140(2):550-9. doi:
905 10.1053/j.gastro.2010.10.057.

906 18. Garg P, Jeppsson S, Dalmasso G, Ghaleb AM, McConnell BB, Yang VW, et al.
907 Notch1 regulates the effects of matrix metalloproteinase-9 on colitis-associated cancer
908 in mice. *Gastroenterology* 2011;141(4):1381-92. doi: 10.1053/j.gastro.2011.06.056.

909 19. Weisshof R, Chermesh I. Micronutrient deficiencies in inflammatory bowel disease.
910 *Curr Opin Clin Nutr Metab Care* 2015;18(6):576-81. doi:
911 10.1097/MCO.0000000000000226.

912 20. Walters JRF, Tasleem AM, Omer OS, Brydon WG, Dew T, le Roux CW. A new
913 mechanism for bile acid diarrhea: defective feedback inhibition of bile acid biosynthesis.
914 *Clin Gastroenterol Hepatol* 2009;7(11):1189-94. doi: 10.1016/j.cgh.2009.04.024.

915 21. Ryan FJ, Ahern AM, Fitzgerald RS, Laserna-Mendieta EJ, Power EM, Clooney AG,
916 et al. Colonic microbiota is associated with inflammation and host epigenomic
917 alterations in inflammatory bowel disease. *Nat Commun* 2020;11(1):1512. doi:
918 10.1038/s41467-020-15342-5.

919 22. Yilmaz B, Juillerat P, Øyås O, Ramon C, Bravo FD, Franc Y, et al. Microbial network
920 disturbances in relapsing refractory Crohn's disease. *Nat Med* 2019;25(2):323-326. doi:
921 10.1038/s41591-018-0308-z.

922 23. Saez-Lara MJ, Gomez-Llorente, Plaza-Diaz J, Gil A. The role of probiotic lactic acid
923 bacteria and bifidobacteria in the prevention and treatment of inflammatory bowel
924 disease and other related diseases: a systematic review of randomized human clinical
925 trials. *Biomed Res Int* 2015;2015:505878. doi: 10.1155/2015/505878.

926 24. Hediger MA, Cléménçon B, Burrier RE, Bruford EA. The ABCs of membrane
927 transporters in health and disease (SLC series): introduction. *Mol Aspects Med*
928 2013;34(2-3):95-107. doi: 10.1016/j.mam.2012.12.009.

929 25. Pérez-Torres S, Iglesias I, Llopis M, Lozano JJ, Antolín M, Guarner F, et al.
930 Transportome Profiling Identifies Profound Alterations in Crohn's Disease Partially
931 Restored by Commensal Bacteria. *J Crohns Colitis* 2016;10(7):850-9. doi:
932 10.1093/ecco-jcc/jjw042.

933 26. Kotka M, Lieden A, Pettersson S, Trinchieri V, Masci A, D'Amato M. Solute carriers
934 (SLC) in inflammatory bowel disease: a potential target of probiotics? *J Clin
935 Gastroenterol* 2008;42 Suppl 3 Pt 1:S133-5. doi: 10.1097/MCG.0b013e31815f5ab6.

936 27. Deleu S, Machiels K, Raes J, Verbeke K, Vermeire S. Short chain fatty acids and its
937 producing organisms: An overlooked therapy for IBD? *EBioMedicine* 2021;66:103293.
938 doi: 10.1016/j.ebiom.2021.103293.

939 28. Salazar N, Neyrinck AM, Bindels LB, Druart C, Ruas-Madiedo P, Cani PD, et al.
940 Functional Effects of EPS-Producing *Bifidobacterium* Administration on Energy
941 Metabolic Alterations of Diet-Induced Obese Mice. *Front Microbiol* 2019;10:1809. doi:
942 10.3389/fmicb.2019.01809.

943 29. Sun S, Luo L, Liang W, Yin Q, Guo J, Rush AM, et al. *Bifidobacterium* alters the gut
944 microbiota and modulates the functional metabolism of T regulatory cells in the context
945 of immune checkpoint blockade. *Proc Natl Acad Sci U S A* 2020;117(44):27509-27515.
946 doi: 10.1073/pnas.1921223117.

947 30. Ray M, Hor PK, Ojha D, Soren JP, Singh SN, Mondal KC. Bifidobacteria and its rice
948 fermented products on diet induced obese mice: analysis of physical status, serum
949 profile and gene expressions. *Benef Microbes* 2018;9(3):441-452. doi:
950 10.3920/BM2017.0056.

951 31. Lim SM, Kim DH. *Bifidobacterium adolescentis* IM38 ameliorates high-fat diet-
952 induced colitis in mice by inhibiting NF-κB activation and lipopolysaccharide production
953 by gut microbiota. *Nutr Res* 2017;41:86-96. doi: 10.1016/j.nutres.2017.04.003.

954 32. Osman N, Adawi D, Molin G, Ahrne S, Berggren A, Jeppsson B. *Bifidobacterium*
955 infantis strains with and without a combination of oligofructose and inulin (OFL)
956 attenuate inflammation in DSS-induced colitis in mice. *BMC Gastroenterol* 2006;6:31.
957 doi: 10.1186/1471-230X-6-31.

958 33. Bloom SM, Bijanki VN, Nava GM, Sun L, Malvin NP, Donermeyer DL, et al.
959 Commensal *Bacteroides* species induce colitis in host-genotype-specific fashion in a
960 mouse model of inflammatory bowel disease. *Cell Host Microbe* 2011;9(5):390-403. doi:
961 10.1016/j.chom.2011.04.009.

962 34. Hooper LV, Wong MH, Thelin A, Hansson L, Falk PG, Gordon JI. Molecular analysis
963 of commensal host-microbial relationships in the intestine. *Science*
964 2001;291(5505):881-4. doi: 10.1126/science.291.5505.881.

965 35. Kelly D, Campbell JI, King TP, Grant G, Jansson EA, Coutts AGP, et al. Commensal
966 anaerobic gut bacteria attenuate inflammation by regulating nuclear-cytoplasmic
967 shuttling of PPAR-gamma and RelA. *Nat Immunol* 2004;5(1):104-12. doi:
968 10.1038/ni1018.

969 36. Wrzosek L, Miquel S, Noordine ML, Bouet S, Chevalier-Curt MJ, Robert V, et al.
970 Bacteroides thetaiotaomicron and Faecalibacterium prausnitzii influence the production
971 of mucus glycans and the development of goblet cells in the colonic epithelium of a
972 gnotobiotic model rodent. *BMC Biol* 2013;11:61. doi: 10.1186/1741-7007-11-61.

973 37. Delday M, Mulder I, Logan ET, Grant G. Bacteroides thetaiotaomicron Ameliorates
974 Colon Inflammation in Preclinical Models of Crohn's Disease. *Inflamm Bowel Dis*
975 2019;25(1):85-96. doi: 10.1093/ibd/izy281.

976 38. Uronis JM, Mühlbauer M, Herfarth HH, Rubinas TC, Jones GS, Jobin C. Modulation
977 of the intestinal microbiota alters colitis-associated colorectal cancer susceptibility.
978 *PLoS One* 2009;4(6):e6026. doi: 10.1371/journal.pone.0006026.

979 39. Hudcovic T, Kozáková H, Kolínská J, Štěpánková R, Hrnčíř T, Tlaskalová-
980 Hogenová H. Monocolonization with Bacteroides ovatus protects immunodeficient SCID
981 mice from mortality in chronic intestinal inflammation caused by long-lasting dextran
982 sodium sulfate treatment. *Physiol Rep* 2009;58(1):101-110. doi:
983 10.33549/physiolres.931340.

984 40. Chang YC, Ching YH, Chiu CC, Liu JY, Hung SW, Huang WC, et al. TLR2 and
985 interleukin-10 are involved in Bacteroides fragilis-mediated prevention of DSS-induced
986 colitis in gnotobiotic mice. *PLoS One* 2017;12(7):e0180025. doi:
987 10.1371/journal.pone.0180025.

988 41. Chu H, Khosravi A, Kusumawardhani IP, Kwon AH, Vasconcelos AC, Cunha LD, et
989 al. Gene-microbiota interactions contribute to the pathogenesis of inflammatory bowel
990 disease. *Science* 2016;352(6289):1116-20. doi: 10.1126/science.aad9948.

991 42. Mazmanian SK, Round JL, Kasper DL. A microbial symbiosis factor prevents
992 intestinal inflammatory disease. *Nature* 2008;453(7195):620-5. doi:
993 10.1038/nature07008.

994 43. Mills RH, Dulai PS, Vázquez-Baeza Y, Saucedo C, Daniel N, Gerner RR, et al.
995 Multi-omics analyses of the ulcerative colitis gut microbiome link *Bacteroides vulgatus*
996 proteases with disease severity. *Nat Microbiol* 2022;7(2):262-276. doi: 10.1038/s41564-
997 021-01050-3.

998 44. Dong G, Chen H, Qi M, Dou Y, Wang Q. Balance between metallothionein and
999 metal response element binding transcription factor 1 is mediated by zinc ions (review).
1000 *Mol Med Rep* 2015;11(3):1582-6. doi: 10.3892/mmr.2014.2969.

1001 45. Tsuji T, Naito Y, Takagi T, Kugai M, Yoriki H, Horie R, et al. Role of metallothioneins
1002 in murine experimental colitis. *Int J Mol Med* 2013;31(5):1037-46. doi:
1003 10.3892/ijmm.2013.1294.

1004 46. Devisscher L, Hindryckx P, Lynes MA, Waeytens A, Cuvelier C, De Vos F, et al.
1005 Role of metallothioneins as danger signals in the pathogenesis of colitis. *J Pathol*
1006 2014;233(1):89-100. doi: 10.1002/path.4330.

1007 47. Waeytens A, De Vos M, Laukens D. Evidence for a potential role of metallothioneins
1008 in inflammatory bowel diseases. *Mediators Inflamm* 2009;2009:729172. doi:
1009 10.1155/2009/729172.

1010 48. Lahiri A, Abraham C. Activation of pattern recognition receptors upregulates
1011 metallothioneins, thereby increasing intracellular accumulation of zinc, autophagy, and
1012 bacterial clearance by macrophages. *Gastroenterology* 2014;147(4):835-46. doi:
1013 10.1053/j.gastro.2014.06.024.

1014 49. Gevers D, Kugathasan S, Denson LA, Vázquez-Baeza Y, Van Treuren W, Ren B, et
1015 al. The treatment-naïve microbiome in new-onset Crohn's disease. *Cell Host Microbe*
1016 2014;15(3):382-392. doi: 10.1016/j.chom.2014.02.005.

1017 50. Dey N, Soergel DA, Repo S, Brenner SE. Association of gut microbiota with post-
1018 operative clinical course in Crohn's disease. *BMC Gastroenterol* 2013;13:131. doi:
1019 10.1186/1471-230X-13-131.

1020 51. Craven M, Egan CE, Dowd SE, McDonough SP, Dogan B, Denkers EY, et al.
1021 Inflammation drives dysbiosis and bacterial invasion in murine models of ileal Crohn's
1022 disease. *PLoS One* 2012;7(7):e41594. doi: 10.1371/journal.pone.0041594.

1023 52. Schaubbeck M, Clavel T, Calasan J, Lagkouvardos I, Haange SB, Jehmlich N, et al.
1024 Dysbiotic gut microbiota causes transmissible Crohn's disease-like ileitis independent of
1025 failure in antimicrobial defence. *Gut* 2016;65(2):225-37. doi: 10.1136/gutjnl-2015-
1026 309333.

1027 53. Kaakoush NO. Insights into the Role of Erysipelotrichaceae in the Human Host.
1028 *Front Cell Infect Microbiol* 2015;5:84. doi: 10.3389/fcimb.2015.00084.

1029 54. van Haaften WT, Blokzijl T, Hofker HS, Olinga P, Dijkstra G, Bank RA, et al.
1030 Intestinal stenosis in Crohn's disease shows a generalized upregulation of genes
1031 involved in collagen metabolism and recognition that could serve as novel anti-fibrotic
1032 drug targets. *Therap Adv Gastroenterol* 2020;13:1756284820952578. doi:
1033 10.1177/1756284820952578.

1034 55. Carstens A, Dicksved J, Nelson R, Lindqvist MM, Andreasson A, Bohr J, et al. The
1035 Gut Microbiota in Collagenous Colitis Shares Characteristics With Inflammatory Bowel
1036 Disease-Associated Dysbiosis. *Clin Transl Gastroenterol* 2019;10(7):e00065. doi:
1037 10.14309/ctg.0000000000000065.

1038 56. Nielsen S, Needham B, Leach ST, Day AS, Jaffe A, Thomas T, et al. Disrupted
1039 progression of the intestinal microbiota with age in children with cystic fibrosis. *Sci Rep*
1040 2016;6:24857. doi: 10.1038/srep24857.

1041 57. Vernocchi P, Del Chierico F, Russo A, Majo F, Rossitto M, Valerio M, et al. Gut
1042 microbiota signatures in cystic fibrosis: Loss of host CFTR function drives the microbiota
1043 enterophenotype. *PLoS One* 2018;13(12):e0208171. doi:
1044 10.1371/journal.pone.0208171.

1045 58. Coffey MJ, Nielsen S, Wemheuer B, Kaakoush NO, Garg M, Needham B, et al. Gut
1046 Microbiota in Children With Cystic Fibrosis: A Taxonomic and Functional Dysbiosis. *Sci*
1047 *Rep* 2019;9(1):18593. doi: 10.1038/s41598-019-55028-7.

1048 59. Cavallari JF, Pokrajac NT, Zlitni S, Foley KP, Henriksbo BD, Schertzer JD. NOD2 in
1049 hepatocytes engages a liver-gut axis to protect against steatosis, fibrosis, and gut
1050 dysbiosis during fatty liver disease in mice. *Am J Physiol Endocrinol Metab*
1051 2020;319(2):E305-E314. doi: 10.1152/ajpendo.00181.2020.

1052 60. Ha CWY, Martin A, Sepich-Poore GD, Shi B, Wang Y, Gouin K, et al. Translocation
1053 of Viable Gut Microbiota to Mesenteric Adipose Drives Formation of Creeping Fat in
1054 Humans. *Cell* 2020;183(3):666-683.e17. doi: 10.1016/j.cell.2020.09.009.

1055 61. Zhuang X, Tian Z, Li N, Mao R, Li X, Zhao M, et al. Gut Microbiota Profiles and
1056 Microbial-Based Therapies in Post-operative Crohn's Disease: A Systematic Review.
1057 *Front Med (Lausanne)* 2021;7:615858. doi: 10.3389/fmed.2020.615858.

1058 62. Frau A, Ijaz UZ, Slater R, Jonkers D, Penders J, Campbell BJ, et al. Inter-kingdom
1059 relationships in Crohn's disease explored using a multi-omics approach. *Gut Microbes*
1060 2021;13(1):1930871. doi: 10.1080/19490976.2021.1930871.

1061 63. Davrandi M, Harris S, Smith PJ, Murray CD, Lowe DM. The Relationship Between
1062 Mucosal Microbiota, Colitis, and Systemic Inflammation in Chronic Granulomatous
1063 Disorder. *J Clin Immunol* 2022;42:312-324. doi: 10.1007/s10875-021-01165-6.

1064 64. Liang JQ, Li T, Nakatsu G, Chen YX, Yau TO, Chu E, et al. A novel faecal
1065 *Lachnospiraceae* marker for the non-invasive diagnosis of colorectal adenoma and
1066 cancer. *Gut* 2020;69(7):1248-1257. doi: 10.1136/gutjnl-2019-318532.

1067 65. Zhao L, Cho WC, Nicolls MR. Colorectal Cancer-Associated Microbiome Patterns
1068 and Signatures. *Front Genet* 2021;12:787176. doi: 10.3389/fgene.2021.787176.

1069 66. Zhou Y, Chen L, Sun G, Li Y, Huang R. Alterations in the gut microbiota of patients
1070 with silica-induced pulmonary fibrosis. *J Occup Med Toxicol* 2019;14:5. doi:
1071 10.1186/s12995-019-0225-1.

1072 67. Ta AD, Ollberding NJ, Karns R, Haberman Y, Alazraki AL, Hercules D, et al.
1073 Association of Baseline Luminal Narrowing With Ileal Microbial Shifts and Gene
1074 Expression Programs and Subsequent Transmural Healing in Pediatric Crohn Disease.
1075 *Inflamm Bowel Dis* 2021;27(11):1707-1718. doi: 10.1093/ibd/izaa339.

1076 68. Liu S, Zhao W, Lan P, Mou X. The microbiome in inflammatory bowel diseases:
1077 from pathogenesis to therapy. *Protein Cell* 2021;12(5):331-345. doi: 10.1007/s13238-
1078 020-00745-3.

1079 69. Hall AB, Yassour M, Sauk J, Garner A, Jiang X, Arthur T, et al. A novel
1080 Ruminococcus gnavus clade enriched in inflammatory bowel disease patients. *Genome*
1081 *Med* 2017;9(1):103. doi: 10.1186/s13073-017-0490-5.

1082 70. Long C, Venema K. Pretreatment of Rapeseed Meal Increases Its Recalcitrant Fiber
1083 Fermentation and Alters the Microbial Community in an *in vitro* Model of Swine Large
1084 Intestine. *Front Microbiol* 2020;11:588264. doi: 10.3389/fmicb.2020.588264.

1085 71. Baxter NT, Schmidt AW, Venkataraman A, Kim KS, Waldron C, Schmidt TM.
1086 Dynamics of Human Gut Microbiota and Short-Chain Fatty Acids in Response to Dietary
1087 Interventions with Three Fermentable Fibers. *mBio* 2019;10(1):e02566-18. doi:
1088 10.1128/mBio.02566-18.

1089 72. Morrison DJ, Preston T. Formation of short chain fatty acids by the gut microbiota
1090 and their impact on human metabolism. *Gut Microbes* 2016;7(3):189-200. doi:
1091 10.1080/19490976.2015.1134082.

1092 73. Wang Y, Gao X, Ghozlane A, Hu H, Li X, Xiao Y, et al. Characteristics of Faecal
1093 Microbiota in Paediatric Crohn's Disease and Their Dynamic Changes During Infliximab
1094 Therapy. *J Crohns Colitis* 2018;12(3):337-346. doi: 10.1093/ecco-jcc/jjx153.

1095 74. Ventin-Holmberg R, Eberl A, Saqib S, Korpela K, Virtanen S, Sipponen T, et al.
1096 Bacterial and Fungal Profiles as Markers of Infliximab Drug Response in Inflammatory
1097 Bowel Disease. *J Crohns Colitis* 2021;15(6):1019-1031. doi: 10.1093/ecco-jcc/jaa252.

1098 75. Dovrolis N, Michalopoulos G, Theodoropoulos GE, Arvanitidis K, Kolios G, Sechi
1099 LA, et al. The Interplay between Mucosal Microbiota Composition and Host Gene-
1100 Expression is Linked with Infliximab Response in Inflammatory Bowel Diseases.
1101 *Microorganisms* 2020;8(3):438. doi: 10.3390/microorganisms8030438.

1102 76. Aden K, Rehman A, Waschina S, Pan WH, Walker A, Lucio M, et al. Metabolic
1103 Functions of Gut Microbes Associate With Efficacy of Tumor Necrosis Factor
1104 Antagonists in Patients With Inflammatory Bowel Diseases. *Gastroenterology*
1105 2019;157(5):1279-1292.e11. doi: 10.1053/j.gastro.2019.07.025.

1106 77. Schierova D, Roubalova R, Kolar M, Stehlikova Z, Rob F, Jackova Z, et al. Fecal
1107 Microbiome Changes and Specific Anti-Bacterial Response in Patients with IBD during
1108 Anti-TNF Therapy. *Cells* 2021;10(11):3188. doi: 10.3390/cells10113188.

1109 78. Nepelska M, de Wouters T, Jacouton E, Béguet-Crespel F, Lapaque N, Doré J, et
1110 al. Commensal gut bacteria modulate phosphorylation-dependent PPAR γ transcriptional
1111 activity in human intestinal epithelial cells. *Sci Rep* 2017;7:43199. doi:
1112 10.1038/srep43199.

1113 79. Byndloss MX, Olsan EE, Rivera-Chávez F, Tiffany CR, Cevallos SA, Lokken KL, et
1114 al. Microbiota-activated PPAR- γ signaling inhibits dysbiotic Enterobacteriaceae
1115 expansion. *Science* 2017;357(6351):570-575. doi: 10.1126/science.aam9949.

1116 80. Abbot EL, McCormack JG, Reynet C, Hassall DG, Buchan KW, Yeaman SJ.
1117 Diverging regulation of pyruvate dehydrogenase kinase isoform gene expression in
1118 cultured human muscle cells. *FEBS J* 2005;272(12):3004-14. doi: 10.1111/j.1742-
1119 4658.2005.04713.x.

1120 81. Bowman TA, O'Keeffe KR, D'Aquila T, Yan QW, Griffin JD, Killion EA, et al. Acyl
1121 CoA synthetase 5 (ACSL5) ablation in mice increases energy expenditure and insulin
1122 sensitivity and delays fat absorption. *Mol Metab* 2016;5(3):210-220. doi:
1123 10.1016/j.molmet.2016.01.001.

1124 82. Kumar J, Rani K, Datt C. Molecular link between dietary fibre, gut microbiota and
1125 health. *Mol Biol Rep* 2020;47(8):6229-6237. doi: 10.1007/s11033-020-05611-3.

1126 83. Ahmadian R, Suh JM, Hah N, Liddle C, Atkins AR, Downes M, et al. PPAR γ
1127 signaling and metabolism: the good, the bad and the future. *Nat Med* 2013;19(5):557-
1128 66. doi: 10.1038/nm.3159.

1129 84. Dubuquoy L, Rousseaux C, Thuru L, Peyrin-Biroulet L, Romano O, Chavatte P, et
1130 al. PPARgamma as a new therapeutic target in inflammatory bowel diseases. *Gut*
1131 2006;55(9):1341-9. doi: 10.1136/gut.2006.093484.

1132 85. Peyrin-Biroulet L, Beisner J, Wang G, Nuding S, Oommen ST, Kelly D, et
1133 al. Peroxisome proliferator-activated receptor gamma activation is required for

1134 maintenance of innate antimicrobial immunity in the colon. *Proc Natl Acad Sci U S A*
1135 2010;107(19):8772-7. doi: 10.1073/pnas.0905745107.

1136 86. Okumura R, Kurakawa T, Nakano T, Kayama H, Kinoshita M, Motooka D, et al.
1137 Lypd8 promotes the segregation of flagellated microbiota and colonic epithelia. *Nature*
1138 2016;532(7597):117-21. doi: 10.1038/nature17406.

1139 87. Burkhardt AM, Tai KP, Flores-Guiterrez JP, Vilches-Cisneros N, Kamdar K,
1140 Barbosa-Quintana O, et al. CXCL17 is a mucosal chemokine elevated in idiopathic
1141 pulmonary fibrosis that exhibits broad antimicrobial activity. *J Immunol*
1142 2012;188(12):6399-406. doi: 10.4049/jimmunol.1102903.

1143 88. Jensen J, Pedersen EE, Galante P, Hald J, Heller RS, Ishibashi M, et al. Control of
1144 endodermal endocrine development by Hes-1. *Nat Genet* 2000;24(1):36-44. doi:
1145 10.1038/71657.

1146 89. Katoh M, Katoh M. Notch signaling in gastrointestinal tract (review). *Int J Oncol*
1147 2007;30(1):247-51.

1148 90. Ma S, Shungin D, Mallick H, Schirmer M, Nguyen LH, Kolde R, et al. Population
1149 Structure Discovery in Meta-Analyzed Microbial Communities and Inflammatory Bowel
1150 Disease. *bioRxiv* 2020; doi: 10.1101/2020.08.31.261214.

1151 91. Limketkai BN, Akobeng AK, Gordon M, Adepoju AA. Probiotics for induction of
1152 remission in Crohn's disease. *Cochrane Database Syst Rev* 2020;7(7):CD006634. doi:
1153 10.1002/14651858.CD006634.pub3.

1154 92. Jakubczyk D, Leszczyńska K, Górska S. The Effectiveness of Probiotics in the
1155 Treatment of Inflammatory Bowel Disease (IBD)- A Critical Review. *Nutrients*
1156 2020;12(7):1973. doi: 10.3390/nu12071973.

1157 93. Rieder F, Fiocchi C, Rogler G. Mechanisms, Management, and Treatment of
1158 Fibrosis in Patients With Inflammatory Bowel Diseases. *Gastroenterology*
1159 2017;152(2):340-350.e6. doi: 10.1053/j.gastro.2016.09.047.

1160 94. Rieder F. The gut microbiome in intestinal fibrosis: environmental protector or
1161 provocateur? *Sci Transl Med* 2013;5(190):190ps10. doi: 10.1126/scitranslmed.3004731.

1162 95. Siva S, Rubin DT, Gulotta G, Wroblewski K, Pekow J. Zinc Deficiency is Associated
1163 with Poor Clinical Outcomes in Patients with Inflammatory Bowel Disease. *Inflamm*
1164 *Bowel Dis* 2017;23(1):152-157. doi: 10.1097/MIB.0000000000000989.

1165 96. Ananthakrishnan AN, Khalili H, Song M, Higuchi LM, Richter JM, Chan AT. Zinc
1166 intake and risk of Crohn's disease and ulcerative colitis: a prospective cohort study. *Int J*
1167 *Epidemiol* 2015;44(6):1995-2005. doi: 10.1093/ije/dyv301.

1168 97. Dai H, Wang L, Li L, Huang Z, Ye L. Metallothionein 1: A New Spotlight on
1169 Inflammatory Diseases. *Front Immunol* 2021;12:739918. doi:
1170 10.3389/fimmu.2021.739918.

1171 98. Jalanka J, Salonen A, Salojarvi J, Ritari J, Immonen O, Marciani L, et al. Effects of
1172 bowel cleansing on the intestinal microbiota. *Gut* 2015;64(10):1562-8. doi:
1173 10.1136/gutjnl-2014-307240.

1174 99. Imhann F, van der Velde KJ, Barbieri R, Alberts R, Voskuil MD, Vich Vila A, et al.
1175 The 1000IBD project: multi-omics data of 1000 inflammatory bowel disease patients;
1176 data release 1. *BMC Gastroenterol* 2019;19(1):5.

1177 100. Heida FH, van Zoonen AGJF, Hulscher JBF, Te Kiefte BJC, Wessels R, Kooi
1178 EMW, et al. A Necrotizing Enterocolitis-Associated Gut Microbiota Is Present in the
1179 Meconium: Results of a Prospective Study. *Clin Infect Dis* 2016;62(7):863-870.

1180 101. Bartram AK, Lynch MDJ, Stearns JC, Moreno-Hagelsieb G, Neufeld JD.
1181 Generation of multimillion-sequence 16S rRNA gene libraries from complex microbial
1182 communities by assembling paired-end illumina reads. *Appl Environ Microbiol*
1183 2011;77(11):3846-52.

1184 102. Bokulich NA, Joseph CM, Allen G, Benson AK, Mills DA. Next-generation
1185 sequencing reveals significant bacterial diversity of botrytized wine. *PLoS One*
1186 2012;e36357.

1187 103. Hu S, Uniken Venema WT, Westra HJ, Vich Vila A, Barbieri R, Voskuil MD, et al.
1188 Inflammation status modulates the effect of host genetic variation on intestinal gene
1189 expression in inflammatory bowel disease. *Nat Commun* 2021;12(1):1122.

1190 104. Ghazi AR, Sucipto K, Rahnavard G, Franzosa EA, McIver LJ, Lloyd-Price J, et al.
1191 High-sensitivity pattern discovery in large, paired multi-omic datasets. *bioRxiv* 2021; doi:
1192 <https://doi.org/10.1101/2021.11.11.468183>

1193 105. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al.
1194 Gene set enrichment analysis: a knowledge-based approach for interpreting genome-
1195 wide expression profiles. *Proc Natl Acad Sci U S A* 2005;102(43):15545-50.

1196 106. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi, Mesirov JP, Tamayo P. The
1197 Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst*
1198 2015;1(6):417-425.

1199 **Acknowledgements**

1200 All authors would like to express their gratitude towards all participants of the 1000IBD
1201 cohort. In addition, the authors would like to thank Kate McIntyre (Scientific Editor,
1202 Department of Genetics, University Medical Center Groningen) for language editing,
1203 and Ren Mao for his kind suggestions on figure styling and structural improvements
1204 (Department of Gastroenterology, The First Affiliated Hospital of Sun Yat-Sen
1205 University, Sun Yat-Sen University, Guangzhou, Guangdong, China).

1206

1207 **Data availability**

1208 The datasets used and/or analyzed for the current study are available from the
1209 corresponding author on reasonable request. The data for the Groningen 1000IBD
1210 cohort can be requested with the accession number EGAS00001002702 (IDs:
1211 EGAD00001003991, EGAD00001008214, and EGAD00001008215).

1212

1213 **Code availability**

1214 All analytic code used for this study can be found at the following link:
1215 https://github.com/GRONINGEN-MICROBIOME-CENTRE/Groningen-Microbiome/tree/master/Projects/IBD_biopsy_project.

1217

1218 **Funding**

1219 RKW is supported by the Seerave foundation and the Netherlands Organization for
1220 Scientific Research. LMS is supported by Takeda. ARB is supported by an MD-PhD
1221 trajectory grant (grant no. 17-57) from the Junior Scientific Masterclass (JSM) of the

1222 University of Groningen the Netherlands. This study was funded by Takeda
1223 Development Center Americas, Inc.

1224

1225 **Conflicts of Interest**

1226 RKW acted as consultant for Takeda, received unrestricted research grants from
1227 Takeda, Johnson & Johnson, Tramedico and Ferring and received speaker fees from
1228 MSD, Abbvie, and Janssen Pharmaceuticals. GD received an unrestricted research
1229 grant from Takeda and speaker fees from Pfizer and Janssen Pharmaceuticals. All
1230 other authors declare no competing interests.

1231

1232 **Authors' contributions**

1233 Conceptualization: RKW. Investigation: SH, ARB, RG, BHJ, RM, AB, IJH, JRB, HJMH,
1234 AVV, LMS and RKW. Methodology: SH, ARB, RG, BHJ, RM, IJH, JRB, HJMH, AVV,
1235 LMS and RKW. Funding acquisition: RKW. Supervision: EAMF, AVV, LMS and RKW.
1236 Writing - original manuscript: SH, ARB and LMS. Writing - review and editing: all
1237 authors.

1238 **Supplementary Methods**

1239 **Polymerase chain reaction (PCR), DNA clean-up, and MiSeq library preparation**
1240 **for mucosal 16 microbiota characterization**

1241 The PCR procedure consisted of the following conditions: an initial cycle of 94°C for 3
1242 min followed by 32 cycles of 94°C for 45 sec, 50°C for 60 sec and 72°C for 90 sec, with
1243 a final extension of 72°C for 10 min. Agarose gel electrophoresis confirmed the
1244 presence of the PCR product (band at ~465 bp) in successfully amplified samples.
1245 Subsequently, DNA samples were thoroughly cleaned by mixing the remainder of the
1246 PCR product with 25 µL Agencourt AMPure XP beads (Beckman Coulter, Brea,
1247 California, USA) followed by an incubation of 5 min at room temperature. Beads were
1248 separated from the mixture by placing the samples within a magnetic bead separator for
1249 2 min. After discarding the cleared solution, beads were washed twice by resuspending
1250 them in 200 µL fresh 80% ethanol, followed by an incubation of 30 sec in the magnetic
1251 bead separator, and again discarding the cleared solution. The pellet was dried for 15
1252 min and resuspended in 52.5 µL 10 mM Tris HCl buffer (pH 8.5). Fifty (50) µL of this
1253 solution was subsequently brought into a new tube. DNA concentrations were
1254 measured using a Qubit® 2.0 fluorometer (Thermo-Fisher Scientific, Waltham,
1255 Massachusetts, USA). To ensure similar library representations across samples, 2 nM
1256 dilutions of each sample were prepared accordingly. A library was created by pooling 5
1257 µL of each diluted sample. Subsequently, 10 µL of the sample pool and 10 µL 0.2 M
1258 NaOH were mixed and incubated for 5 min to allow denaturation of the sample DNA.
1259 980 µL of the HT1 buffer of the MiSeq 2x300 cartridge was then added to this mixture.
1260 Next, a denatured diluted PhiX solution was created by combining 2 µL 10 nM PhiX
1261 library with 3 µL 10 mM Tris HCl buffer (pH 8.5) with 0.1% Tween-20. 5 µL 0.2 M NaOH
1262 was added to this mixture and incubated for 5 min at room temperature. This 10 µL
1263 mixture was eventually mixed with 990 µL HT1 buffer. From the diluted sample pool,
1264 150 µL was combined with 50 µL of the diluted PhiX solution, which was further diluted
1265 by the addition of 800 µL HT1 buffer. Finally, 600 µL of the prepared library solution was
1266 loaded into the sample loading reservoir of the 2x300 MiSeq cartridge for 16S rRNA
1267 amplicons sequencing (MiSeq Benchtop Sequencer, Illumina, San Diego, California,

1268 USA). Samples with low DNA concentrations after clean-up (quality score < 0.9) were
1269 discarded by PANDAseq to increase quality of sequence read-outs.

1270 **Supplementary Table S1.** Nucleotide sequences of primers used for library
1271 construction for bacterial 16S rRNA gene (Illumina) sequencing.

Primer	Sequence
V3_F_modified	aatgatacggcgaccaccgagat <u>ctacactcttccctacacgacgcttccgatctNNNNCCT</u> ACGGGAGGCAGCAG
V4_1R	caagcagaagacggatacgagat <u>CGTGATgtgactggagttcagacgtgtgctctccgatct</u> GGACTACHVGGGTWTCTAAT
V4_2R	caagcagaagacggatacgagat <u>ACATCGgtgactggagttcagacgtgtgctctccgatct</u> GGACTACHVGGGTWTCTAAT
V4_3R	caagcagaagacggatacgagat <u>GCCTAAgtgactggagttcagacgtgtgctctccgatct</u> GGACTACHVGGGTWTCTAAT
V4_4R	caagcagaagacggatacgagat <u>TGGTCAgtgactggagttcagacgtgtgctctccgatct</u> GGACTACHVGGGTWTCTAAT
V4_5R	caagcagaagacggatacgagat <u>CACTGTgtgactggagttcagacgtgtgctctccgatct</u> GGACTACHVGGGTWTCTAAT
V4_6R	caagcagaagacggatacgagat <u>ATTGGCgtgactggagttcagacgtgtgctctccgatct</u> GGACTACHVGGGTWTCTAAT
V4_7R	caagcagaagacggatacgagat <u>GATCTGgtgactggagttcagacgtgtgctctccgatct</u> GGACTACHVGGGTWTCTAAT

V4_8R	caagcagaagacggcatacgagat TCAAGT tgactggagttcagacgtgtgctttccgatct GGACTACHVGGGTWTCTAAT
V4_9R	caagcagaagacggcatacgagat CTGATC tgactggagttcagacgtgtgctttccgatct GGACTACHVGGGTWTCTAAT
V4_10R	caagcagaagacggcatacgagat AAGCTA tgactggagttcagacgtgtgctttccgatct GGACTACHVGGGTWTCTAAT
V4_11R	caagcagaagacggcatacgagat GTAGCC tgactggagttcagacgtgtgctttccgatct GGACTACHVGGGTWTCTAAT
V4_12R	caagcagaagacggcatacgagat TACAAG tgactggagttcagacgtgtgctttccgatct GGACTACHVGGGTWTCTAAT
V4_13R	caagcagaagacggcatacgagat CGTACT tgactggagttcagacgtgtgctttccgatct GGACTACHVGGGTWTCTAAT
V4_14R	caagcagaagacggcatacgagat GACTGA tgactggagttcagacgtgtgctttccgatct GGACTACHVGGGTWTCTAAT
V4_15R	caagcagaagacggcatacgagat GCTCAA tgactggagttcagacgtgtgctttccgatct GGACTACHVGGGTWTCTAAT
V4_16R	caagcagaagacggcatacgagat TCGCTT tgactggagttcagacgtgtgctttccgatct GGACTACHVGGGTWTCTAAT

1272 Bold uppercase letter highlight the index sequences. Lowercase letters indicate adapter sequences
1273 necessary for flow-cell binding. Underlined lowercase letters indicate binding sites for Illumina sequencing
1274 primers. Regular uppercase letters indicate the V3 and V4 region primers (341F for the forward primer,
1275 806R for the reverse primer). The inclusion of four maximally degenerated bases ("NNNN") maximizes
1276 the diversity during the first four bases of the run, which is important for unique cluster identification and
1277 base-calling accuracy.

1279 **Supplementary Table Index**

1280 *All Supplementary Tables have been uploaded separately for peer-review.*

1281 **Supplementary Table S1.** Differential gene expression analyses between non-inflamed
1282 and inflamed biopsies from ileal CD (group 1), colonic CD (group 2) and UC (group 3).

1283 **Supplementary Table S2.** Differential gene expression analysis between inflamed
1284 biopsies from patients with CD (reference) and patients with UC.

1285 **Supplementary Table S3.** Differential expression analysis of deconvoluted cell types
1286 between inflamed colonic biopsies of patients with CD (reference) and patients with UC.

1287 **Supplementary Table S4.** Relative abundances of mucosal bacterial groups in different
1288 groups (CD, UC and non-IBD controls) and biopsy locations (ileum or colon).

1289 **Supplementary Table S5.** Comparison of relative abundances of bacterial groups
1290 between non-IBD controls and ileal CD (group 1), colonic CD (group 2) and UC (group
1291 3).

1292 **Supplementary Table S6.** Hierarchical analysis performed using an end-to-end
1293 statistical algorithm (HALIA) demonstrating the main associations between mucosal
1294 bacterial groups and clinical phenotypes.

1295 **Supplementary Table S7.** Genes and bacteria contained in component pair 1 from
1296 sparse-CCA analysis (FDR<0.05).

1297 **Supplementary Table S8.** Pathway annotation of genes involved in component pair 1
1298 from sparse-CCA analysis (FDR<0.05).

1299 **Supplementary Table S9.** Genes and bacteria contained in component pair 3 from
1300 sparse-CCA analysis (FDR<0.05).

1301 **Supplementary Table S10.** Pathway annotation of genes involved in component pair 3
1302 from sparse-CCA analysis (FDR<0.05).

1303 **Supplementary Table S11.** Genes and bacteria contained in component pair 7 from
1304 sparse-CCA analysis (FDR<0.05).

1305 **Supplementary Table S12.** Pathway annotation of genes involved in component pair 7
1306 from sparse-CCA analysis (FDR<0.05).

1307 **Supplementary Table S13.** Genes and bacteria contained in component pair 8 from
1308 sparse-CCA analysis (FDR<0.05).

1309 **Supplementary Table S14.** Pathway annotation of genes involved in component pair 8
1310 from sparse-CCA analysis (FDR<0.05).

1311 **Supplementary Table S15.** Genes and bacteria contained in component pair 5 from
1312 sparse-CCA analysis (FDR<0.05).

1313 **Supplementary Table S16.** Pathway annotation of genes involved in component pair 5
1314 from sparse-CCA analysis (FDR<0.05).

1315 **Supplementary Table S17.** Genes and bacteria contained in component pair 9 from
1316 sparse-CCA analysis (FDR<0.05).

1317 **Supplementary Table S18.** Pathway annotation of genes involved in component pair 9
1318 from sparse-CCA analysis (FDR<0.05).

1319 **Supplementary Table S19.** Individual pairwise gene-bacteria associations.

1320 **Supplementary Table S20.** Genes and bacteria associated with fibrostenotic
1321 CD/Montreal B2 (reference: Montreal B1).

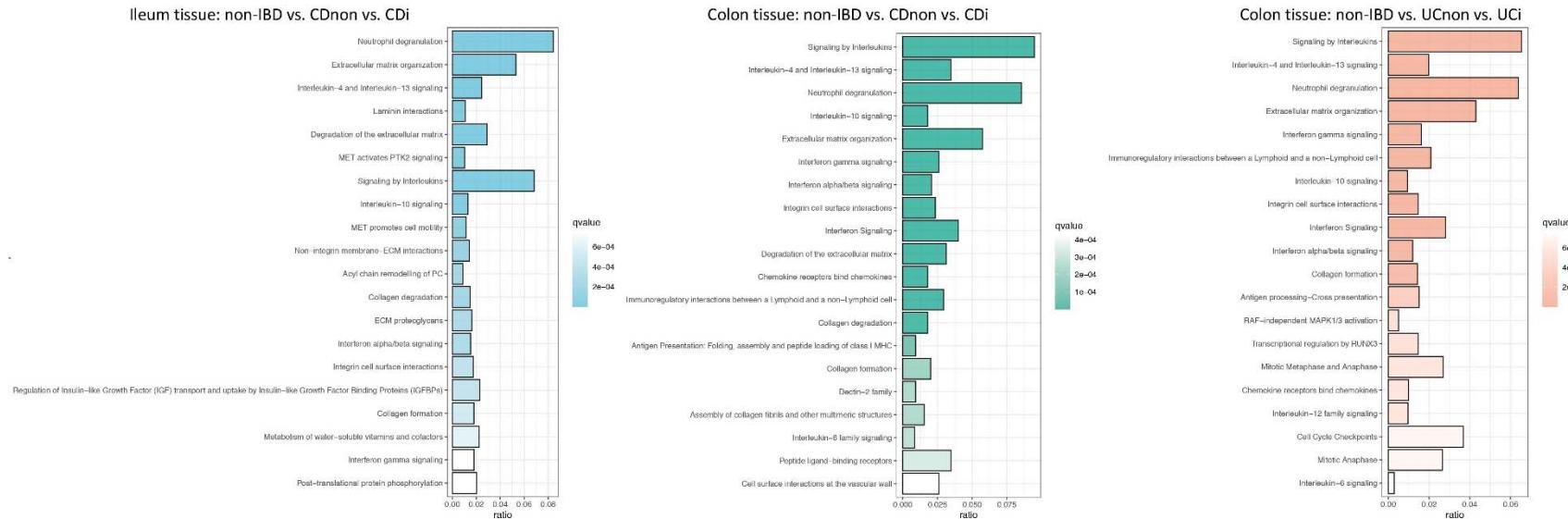
1322 **Supplementary Table S21.** Microbiota-associated gene clusters in patients with non-
1323 stricturing, non-penetrating disease (Montreal B1) and fibrostenotic CD (Montreal B2)
1324 including microbiota-associated pathway annotation and cluster comparisons.

1325 **Supplementary Table S22.** Genes and bacteria associated with TNF- α -antagonists
1326 use (reference: non-users).

1327 **Supplementary Table S23.** Microbiota-associated gene clusters in patients not using
1328 and using TNF- α -antagonists including microbiota-associated pathway annotation and
1329 cluster comparisons.

1330 **Supplementary Table S24.** Individual pairwise gene-bacteria associations and their
1331 interaction with the degree of mucosal dysbiosis (Lloyd-Price *et al.*, *Nature* 2019).

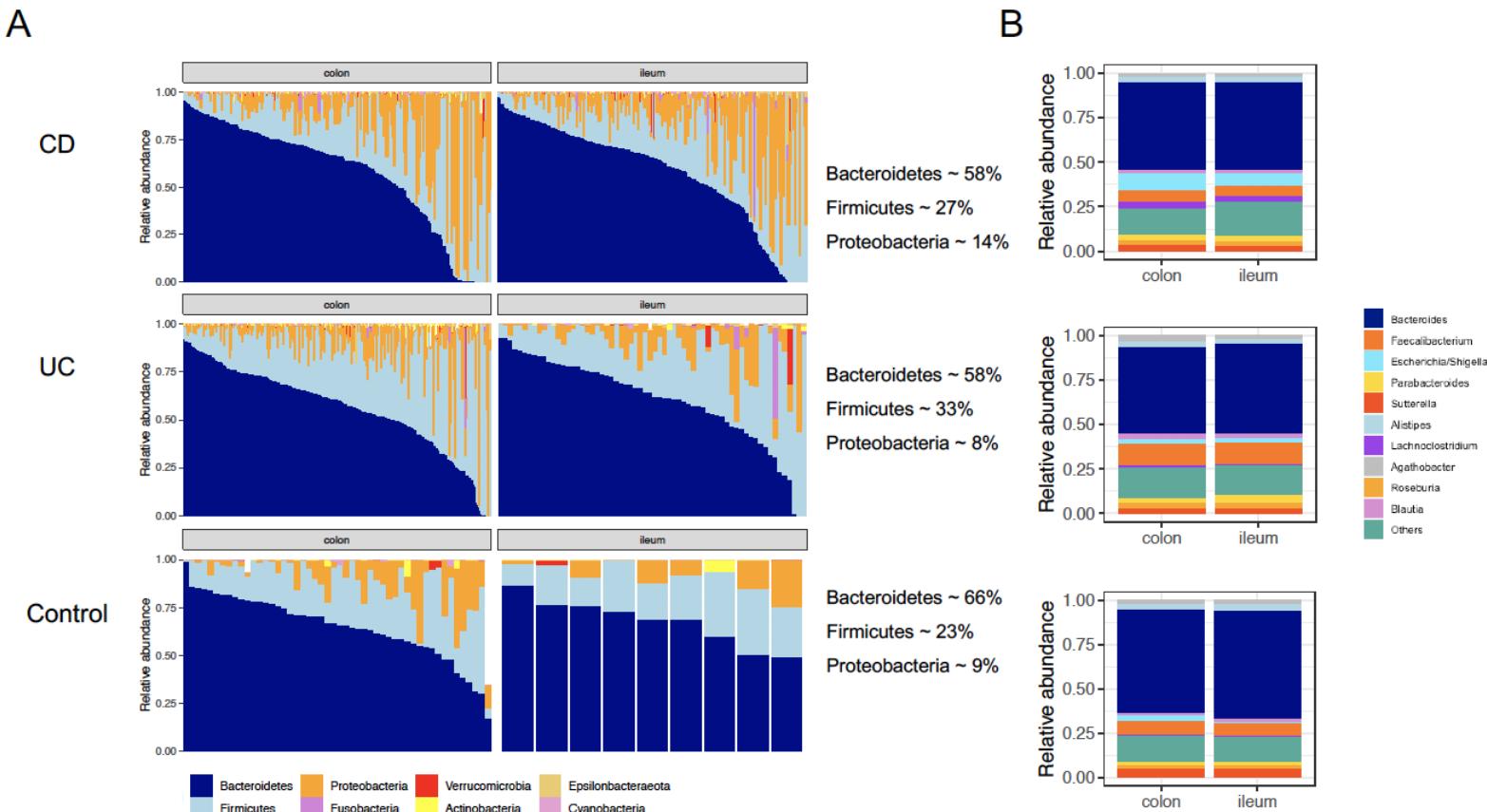
1332 **Supplementary Table S25.** Mucosal microbiota and other phenotypic factors
1333 explaining variation in mucosal cell type enrichment in patients with IBD.

1334 **Extended Data Figures**1335 **Extended Data Fig. S1**

1336

1337 **Extended Data Fig. S1. Analysis of pathways associated with each comparative gene expression analysis.** The main pathways
 1338 associated with inflamed ileal tissue in patients with CD (blue) include neutrophil degranulation, extracellular matrix (ECM) organization
 1339 and IL-4/IL-13-signaling. Similar pathways were overexpressed in inflamed colonic tissue from patients with CD (green), but with a more
 1340 prominent contribution from interleukin signaling pathways. Interleukin signaling pathways were also dominantly expressed in inflamed
 1341 colonic tissue from patients with UC (orange), with other pathways expressed including neutrophil degranulation, ECM pathways,
 1342 interferon gamma signaling and immunoregulatory interactions between lymphoid and non-lymphoid cells. Pathways were annotated
 1343 using the Reactome pathway database. Abbreviations: CDi, inflamed tissue from patients with Crohn's disease; CD-non, non-inflamed
 1344 tissue from patients with Crohn's disease; UCi, inflamed tissue from patients with ulcerative colitis; UC-non, non-inflamed tissue from
 1345 patients with ulcerative colitis.

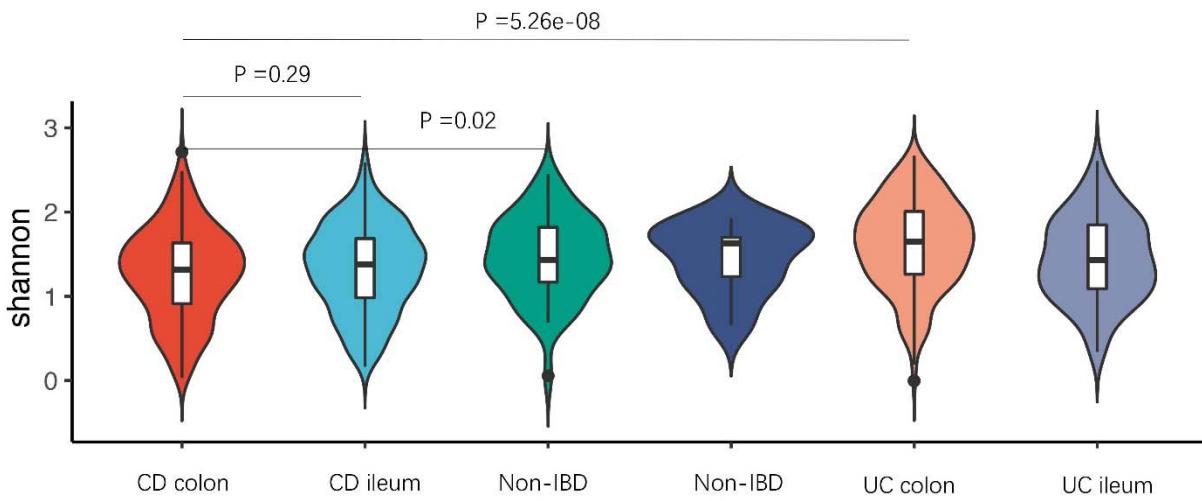
1346 Extended Data Fig. S2



1347

1348 **Extended Data Fig. S2. Mucosal 16S rRNA gene sequencing characterization demonstrates distinct compositional differences in relative**
1349 **abundances on (A) bacterial phylum level and (B) bacterial genus level. Abbreviations: CD, Crohn's disease; UC, ulcerative colitis.**

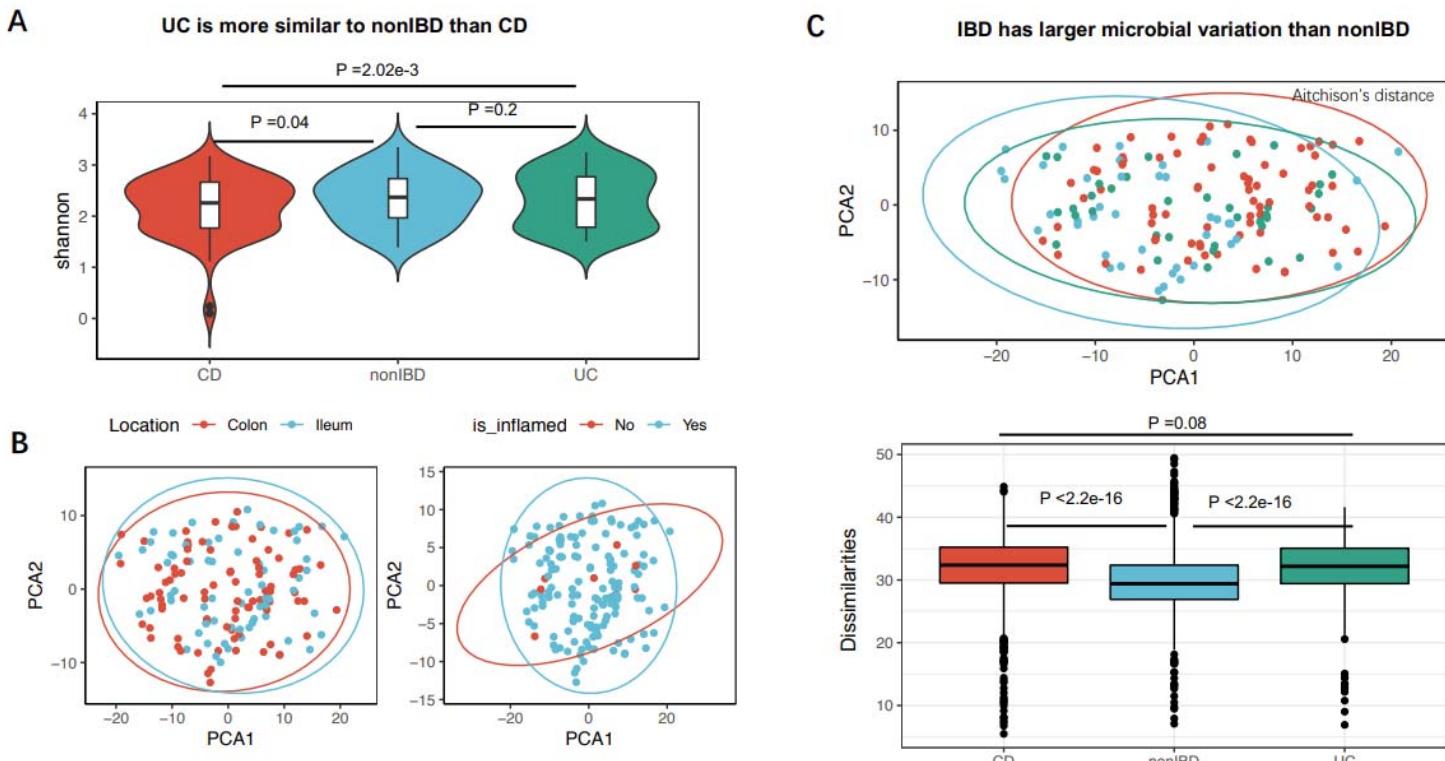
1350

1351 **Extended Data Fig. S3**

1352

1353 **Extended Data Fig. S3. Microbial alpha-diversity (Shannon index) is significantly lower in colonic biopsies from patients with CD**
 1354 **compared to colonic biopsies derived from patients with UC or controls.** This indicates that this difference is not solely attributable to ileal
 1355 biopsies from patients with CD.

1356

1357 **Extended Data Fig. S4**

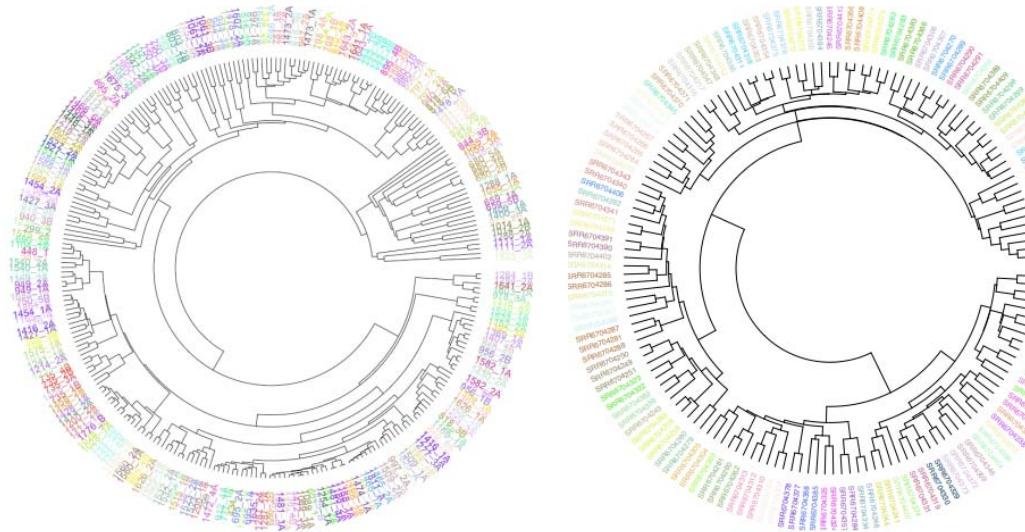
1358

1359 Extended Data Fig. S4. Replication of overall mucosal microbiota characterization in patients with IBD and non-IBD controls.

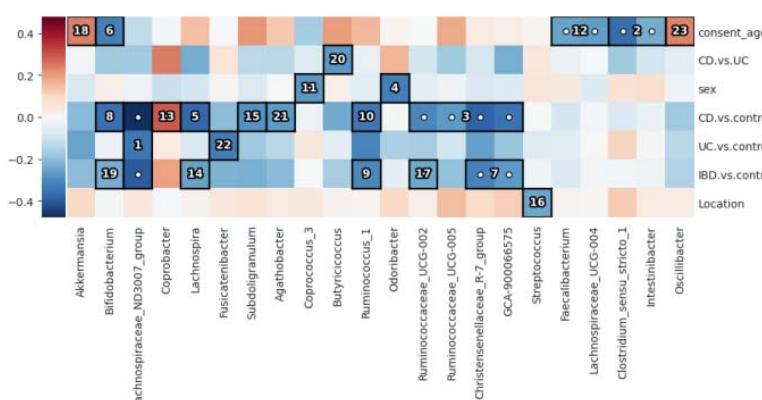
1360 Replication was performed in data derived from the HMP2 cohort study [13]. **a**, Microbial alpha-diversity (Shannon index) was lowest in
1361 patients with CD (n=85) compared to patients with UC (n=46) and non-IBD controls (n=45). **b**, PCA plots based on Aitchison's distances
1362 and stratified by tissue location and inflammatory status (colors as in **a**). **c**, PCA plot showing microbial dissimilarity (Aitchison's distances)
1363 in CD, UC and non-IBD controls. **d**, Microbial dissimilarity is highest in samples from patients with CD, followed by patients with UC and
1364 non-IBD controls. CD, Crohn's disease; PCA, principal component analysis; UC, ulcerative colitis.

1365 **Extended Data Fig. S5**

A



B

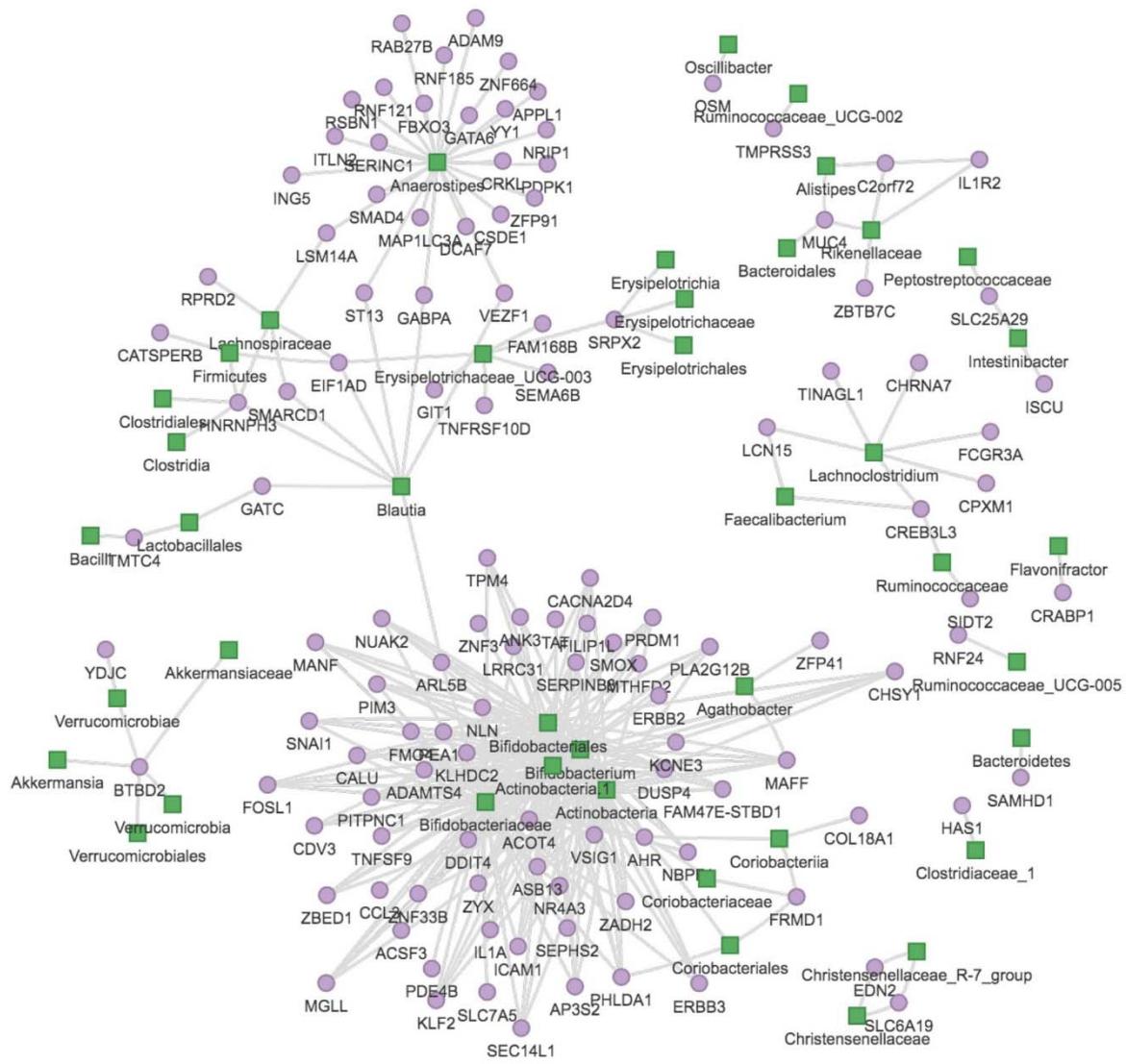


1366

1367 **Extended Data Fig. S5. Composition of the mucosal microbiota is highly personalized and**
1368 **influenced by disease parameters and clinical factors in patients with IBD and controls. (A)**
1369 Hierarchical clustering analysis demonstrating that tissue samples from the same individual
1370 (paired samples) clearly cluster together (colors indicate unique individuals). (B) Hierarchical
1371 analysis performed using an end-to-end statistical algorithm (HAIIA) showing the main phenotypic
1372 factors that correlate with intestinal mucosal microbiota composition. Heatmap color palette
1373 indicates normalized mutual information. Numbers and dots in cells identify the significant pairs of
1374 features (phenotypic factors vs. bacterial taxa) in patients with IBD and controls. Abbreviations:
1375 BMI, body-mass index; CD, Crohn's disease; UC, ulcerative colitis.

1376

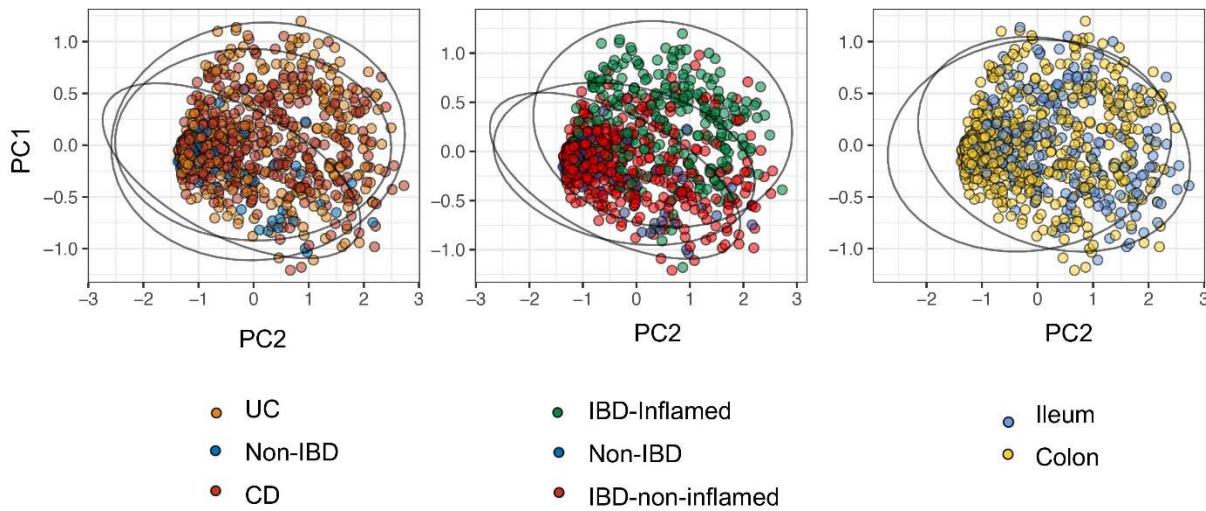
1377 **Extended Data Fig. S6**



1378

1379 **Extended Data Fig. S6. Network graph displaying significant individual gene–bacteria interactions.**
 1380 Green squares indicate bacterial groups. Purple dots indicate host gene expression. Each connecting line
 1381 indicates statistically significant gene–bacteria associations after adjustment for age, sex, batch,
 1382 medication use, tissue inflammatory status and tissue location. Most individual gene–bacteria
 1383 associations (94%) overlap with the results from the sparse-CCA analysis (**Figure 4**).

1384 **Extended Data Fig. S7**



1385

1386 **Extended Data Fig. S7. Principal component analysis (PCA) plots demonstrating variation in cell**
1387 **type-enrichment labeled by diagnosis, biopsy inflammatory status and intestinal location.** Each
1388 dot represents one tissue sample. Left: Patients with IBD, both CD and UC, show significantly different
1389 intestinal cell type composition compared to controls. Middle: Tissue inflammatory status induces shifts in
1390 cell type composition, showing differences between non-inflamed IBD tissue vs. control tissue and
1391 inflamed IBD tissue vs. control tissue. Right: Tissue location (ileum vs. colon) also demonstrates distinct
1392 variation in cell type composition.

1393 **Supplementary Results**

1394 **Box 1. Individual mucosal gene–bacteria associations and their potential**
1395 **biological implications in IBD.**

1396 *Mucosal bifidobacteria positively associate with aryl hydrocarbon receptor (AHR) and*
1397 *ABC-transporter (ABCC1) expression levels*

1398 The positive association between *AHR* expression and bifidobacteria could be
1399 explained by the fact that *Bifidobacterium* spp. can produce aromatic lactic acids such
1400 as indole-3-lactic acid (out of aromatic amino acids like tryptophan) via aromatic lactate
1401 dehydrogenase, which in turn activates the host aryl hydrocarbon receptor [1,2].
1402 Activation of the aryl hydrocarbon receptor, a crucial regulator of intestinal homeostasis
1403 and immune responses, leads to a reduction of inflammation in intestinal epithelial cells
1404 [3] and confers immunoprotective effects [4].

1405 Another intriguing observation is the positive association between bifidobacteria and
1406 host expression of the *ABCC1* gene. *ABCC1* is a member of the ATP-binding cassette
1407 transporters (ABC transporters, and also known as multidrug resistance-associated
1408 protein 1, MRP1) that has multiple physiological functions, but it may also confer
1409 pathophysiological sequelae, especially in the context of cancer [5]. Under physiological
1410 circumstances, it detoxifies endogenously generated toxic substances (as well as
1411 xenobiotics), protects against oxidative stress, transports leukotrienes and lipids and
1412 may facilitate the cellular export and body distribution of vitamin B₁₂ [6]. Interestingly,
1413 several *Bifidobacterium* species (e.g. *B. animalis*, *B. longum* and *B. infantis*) can
1414 synthesize vitamin B₁₂, which is subsequently absorbed in the large intestine via
1415 unknown mechanisms [7-9].

1416 *Mucosal bifidobacteria associate with FOSL1, a subunit of the AP-1 transcription factor*
1417 Associations between mucosal *Bifidobacterium* bacteria and expression of *FOSL1*
1418 genes were amongst the top significant individual gene–bacteria interactions. Fos-
1419 related antigen 1 (FRA1), encoded by *FOSL1*, is a subunit of the activator protein 1
1420 (AP-1) transcription factor. In the intestine, the AP-1 transcription factor is commonly

1421 activated in response to inflammatory stimuli and has been implicated in IBD [10]. More
1422 specifically, an interaction may exist between AP-1 activity and the glucocorticoid
1423 receptor, which may be part of the anti-inflammatory effects of steroid treatment [11]. In
1424 steroid-resistant patients with CD, AP-1 activation was primarily observed in the nuclei
1425 of intestinal epithelial cells, whereas this activation was restricted to lamina propria
1426 macrophages in steroid-sensitive patients [10]. This suggests a differing cellular
1427 activation pattern of AP-1 activation in steroid-resistant patients where the expression of
1428 this transcription factor may interfere with the activity of the glucocorticoid response. In
1429 an experimental study in which pregnant mice were supplemented with butyrate, FOS
1430 genes, including *Fosl1*, were observed to be downregulated in the colon and associated
1431 with protection against experimentally-induced colitis [12]. Although there are currently
1432 no reports of potential immune-modulating effects for *Fosl1*, it has 85% homology with
1433 *Fosl2*, another AP-1 transcription factor. A recent study demonstrated that *Fosl2* is
1434 important in T-reg development and control of autoimmunity. Interestingly, several
1435 GWASs have reported associations of a SNP located in the promoter region of *Fosl2*
1436 with IBD [13-15], and the presence of this SNP was also shown to correlate with *Fosl2*
1437 expression in blood cells of patients with IBD [16]. In the context of T-regs, *Fosl2* also
1438 appears to be important as it is a determinant of a highly suppressive subpopulation of
1439 T-regs in humans that are particularly enriched in the lamina propria of patients with CD,
1440 supporting wound healing in the intestinal mucosa [17]. Although speculative,
1441 bifidobacteria, as well as their metabolites such as butyrate, may potentially confer
1442 immune-modulating properties via interaction with *Fosl1* expression.

1443 *Mucosal bifidobacteria positively associate with Krüppel-like factor 2 (KLF2) expression*
1444 Krüppel-like factor 2 (encoded by *KLF2*) is a negative regulator of intestinal
1445 inflammation, and its expression is found to be reduced in patients with IBD [18]. *KLF2*
1446 also negatively regulates differentiation of adipocytes and strongly inhibits PPAR-γ
1447 expression, which prevents differentiation of preadipocytes into adipocytes and thereby
1448 prevents adipogenesis [19]. *KLF2* also plays an important role in endothelial physiology,
1449 where it may act as a molecular switch by regulating endothelial cell function in
1450 inflammatory disease states [20]. Interestingly, *KLF2* modifies the trafficking of T-regs,

1451 as increased *KLF2* expression in T-reg cells promotes the induction of peripheral
1452 immunological tolerance, whereas, in the absence of its expression, T-reg cells are unable
1453 to effectively migrate to secondary lymphoid tissues [21]. Indeed, it was demonstrated
1454 in mouse experiments that mice developed IBD in the presence of *KLF2*-deficient T-
1455 reg cells, which were unable to prevent colitis by disrupted co-trafficking of effector and
1456 regulatory T cells. In light of these considerations, mucosal bifidobacteria may confer
1457 beneficial immune-modulating properties by upregulating *KLF2* expression, thereby
1458 stimulating T-reg migration and contributing to immunological self-tolerance in the
1459 context of IBD.

1460 *Mucosal Anaerostipes bacteria positively associate with host SMAD4 expression*
1461 *Anaerostipes*, which belong to the *Lachnospiraceae* family, are anaerobic bacteria that
1462 are well-known butyrate-producers. Butyrate serves as the primary energy source for
1463 colonic epithelial cells and is characterized by anti-inflammatory and anti-carcinogenic
1464 properties. *SMAD4* is an important intracellular effector of the TGF- β superfamily of
1465 proteins. These proteins have important functions in alleviating intestinal inflammation
1466 and maintenance of gut mucosal homeostasis. Haploinsufficiency of *SMAD4* in mice
1467 and humans has been associated with an increased susceptibility to colonic
1468 inflammation [22]. In patients with CD, reduced epithelial protein levels of *SMAD4* were
1469 observed that was associated with disease activity, indicating defective mucosal TGF- β
1470 signaling during active intestinal inflammation. In an experimental animal study, mice
1471 with an epithelial deletion of *Smad4* presented with macroscopic invasive
1472 adenocarcinoma of the distal colon and rectum 3 months after DSS-induced colitis [23].
1473 Indeed, *SMAD4* mutations in humans are linked to juvenile polyposis syndrome and
1474 associated with poor disease outcome in several types of cancer [24-27]. Using RNA-
1475 seq analysis, a strong inflammatory expression profile was observed after *SMAD4*
1476 deletion, with expression of various inflammatory cytokines and chemokines, including
1477 CCL20. In addition, it was demonstrated that CCL20 could be repressed by *SMAD4* in
1478 colonic epithelial cells, proving that TGF- β signaling could block the induction of CCL20
1479 expression to protect against the development of colitis-associated cancer.

1480 In an experimental study involving human hepatic stellate cells, butyrate was
1481 demonstrated to be protective against diet-induced nonalcoholic steatohepatitis and
1482 liver fibrosis via suppression of TGF- β signaling pathways in which SMAD proteins are
1483 involved. Although butyrate mainly showed antifibrotic effects via reduction of non-
1484 canonical TGF- β signaling cascades, there was also a significant increase in the
1485 expression of SMAD4 with the addition of butyrate on top of TGF- β treatment [28]. We
1486 found *Anaerostipes* bacteria to also be strongly associated with expression of *ZNF644*,
1487 a zinc finger protein that is positively regulated by intracellular zinc concentrations.
1488 Depletion of intracellular zinc levels, or even zinc deficiency, may have destabilizing
1489 effects on SMAD proteins and thereby impair the TGF- β signaling pathway [29].

1490 *Mucosal Verrucomicrobia bacteria inversely associate with expression of the IBD*
1491 *susceptibility gene YDJC*

1492 We observed significant inverse associations between Verrucomicrobia bacteria, of
1493 which *Akkermansia muciniphila* is a well-known member, and the expression of the
1494 *YDJC* gene, which encodes for the YdjC chitooligosaccharide deacetylase homolog
1495 (YdjC) protein. This gene has been identified as a shared susceptibility gene for CD, UC
1496 and psoriasis [13,30,31]. *YDJC* was originally identified as a celiac disease–associated
1497 susceptibility locus, but some SNPs were also associated with CD as well as with
1498 pediatric-onset CD [32]. YdjC catalyzes the deacetylation of acetylated carbohydrates,
1499 an important reaction in the degradation of oligosaccharides [33]. *YDJC* expression has
1500 been associated with tumor progression in studies of lung cancer [34,35]. The observed
1501 inverse association between *Akkermansia* and *YDJC* expression may suggest a
1502 potential protective role of *Akkermansia*, as decreased *YDJC* expression may mitigate
1503 its pro-carcinogenic effects. Despite the association between *YDJC* and the
1504 susceptibility to IBD on a genetic level, its precise functional role remains largely
1505 unknown [32].

1506 *Mucosal Alistipes bacteria positively associate with MUC4 expression*

1507 The bacterial genus *Alistipes*, belonging to family *Rikenellaceae* and phylum
1508 Bacteroidetes, is a recently discovered bacterial species, of which many have been
1509 isolated from the human gut microbiome. The role of *Alistipes* in health and disease is

1510 still unclear. Some evidence indicates that it may confer protective effects to the host,
1511 but other studies report pathogenic effects, e.g. in colorectal cancer development. A key
1512 factor believed to determine the relative abundance of *Alistipes* is the dysbiotic state of
1513 the gut microbiome [36]. In IBD, there is also conflicting data about the pathogenicity of
1514 *Alistipes* species. *Alistipes finegoldii* has been demonstrated to exert anti-inflammatory
1515 effects in experimental models of colitis [37]. Likewise, another study found an
1516 increased abundance of *Alistipes* in *NOD2*-knockout mice that had less severe (TNBS-
1517 induced) colitis compared to wild-type mice [38]. It has also been reported that *Alistipes*
1518 abundance could increase after taking probiotic supplements, which in turn may protect
1519 against hepatocellular cancer growth in an experimental setting [39]. However,
1520 metagenomic studies have shown that *Alistipes* abundances were increased in mouse
1521 models of spontaneous CD-like ileitis terminalis as compared to wild-type mice,
1522 suggesting that *Alistipes* species may also play a pathogenic role by eliciting segmental
1523 ileitis [40,41].

1524 *MUC4* encodes for mucin 4, a protein found in the glycocalyx present on the intestinal
1525 epithelium. Deletion or knockouts of *Muc4* have demonstrated protective effects in
1526 mouse models, as shown by lower levels of proinflammatory factors and resistance
1527 against DSS-induced colitis. It is still unclear how this protective mechanism of *MUC4*
1528 deletion works, but it has been hypothesized that it may trigger the concomitant
1529 upregulation of other mucin proteins (e.g. *MUC1-3*) as these genes have been observed
1530 to be highly expressed in *Muc4*-knockout mice with DSS-induced colitis [42,43]. Based
1531 on this, we speculate that the positive association between *Alistipes* abundance and
1532 *MUC4* expression may imply a potential pathogenic role of *Alistipes* in the context of
1533 IBD-associated dysbiosis. However, in our data, we did not observe a significant
1534 interaction via dysbiotic status between *Alistipes* abundance and *MUC4* expression.

1535 *Mucosal Oscillibacter bacteria positively associate with OSM expression*

1536 *Oscillibacter*-like bacteria, which include *Oscillibacter* and *Oscillospira*, are commonly
1537 detected in human gut microbial communities, although their exact physiological role is
1538 not fully understood. Previously, it was reported that *Oscillibacter* may be a potentially
1539 important bacterium in mediating high fat diet-induced intestinal dysfunction, which was

1540 supported by a negative association between *Oscillibacter* and intestinal barrier function
1541 parameters [44]. Similarly, the abundance of *Oscillibacter* has been reported as a key
1542 bacterial group associated with colitis development in DSS-induced colitis in mice and
1543 with prenatal stress in rodents [45,46]. However, a recent study linking gut microbiota
1544 profiles to sulfur metabolism in patients with CD demonstrated that *Oscillibacter*
1545 abundance was enriched in patients with inactive compared to active disease but
1546 diminished in patients with IBD compared to controls [47,48]. Thus, similar to
1547 *Bacteroides* and *Alistipes*, the exact functional role of *Oscillibacter* in the context of IBD
1548 remains elusive, but it will likely depend on gut microbial dysbiosis and the intestinal
1549 (inflammatory) environment. The OSM gene encodes for the oncostatin M protein, a
1550 well-known inflammatory mediator in IBD that drives intestinal inflammation, mainly *via*
1551 activation of JAK-STAT and PI3K-Akt pathways [49]. Besides induction of other
1552 inflammatory events, it primarily triggers the production of various cytokines,
1553 chemokines and adhesion molecules that contribute to intestinal inflammation [50]. In
1554 addition, OSM is a marker for non-responsiveness to TNF- α -antagonists in patients with
1555 IBD [51]. Considering these findings, the positive association between OSM expression
1556 and *Oscillibacter* abundance we observe supports a potentially pathogenic role for this
1557 bacterial species in IBD.

1558

1559 *Mucosal Blautia bacteria associate with host ST13 expression levels*
1560 Hsc70-interacting protein, encoded by the *ST13* gene, mediates the assembly of the
1561 human glucocorticoid receptor, which requires involvement of intracellular chaperone
1562 proteins such as heat shock proteins HSP70 and HSP90 [52]. Reduced expression of
1563 *ST13* has been observed in patients with colorectal cancer, suggesting that *ST13* may
1564 constitute a candidate tumor-suppressor gene [53,54]. The positive association we
1565 observe between mucosal *Blautia* abundance and *ST13* gene expression may therefore
1566 point to a protective anti-carcinogenic role for *Blautia* in the intestines.

1567

1568 **Supplementary References to Box 1**

1569 1. Wong CB, Tanaka A, Kuhara T, Xiao JZ. Potential effects of Indole-3-Lactic Acid, a
1570 Metabolite of Human Bifidobacteria, on NGF-induced Neurite Outgrowth in PC12 Cells.
1571 *Microorganisms* 2020;8(3):398. doi: 10.3390/microorganisms8030398.

1572 2. Laursen MF, Sakanaka M, von Burg N, Mörbe U, Andersen D, Moll JM, et al.
1573 Bifidobacterium species associated with breastfeeding produce aromatic lactic acids in
1574 the infant gut. *Nat Microbiol* 2021;6(11):1367-1382. doi: 10.1038/s41564-021-00970-4.

1575 3. Ehrlich AM, Pacheco AR, Henrick BM, Taft D, Xu G, Huda MN, et al. Indole-3-lactic
1576 acid associated with Bifidobacterium-dominated microbiota significantly decreases
1577 inflammation in intestinal epithelial cells. *BMC Microbiol* 2020;20(1):357. doi:
1578 10.1186/s12866-020-02023-y.

1579 4. Henrick BM, Rodriguez L, Lakshmikanth T, Pou C, Henckel E, Arzoomand A, et al.
1580 Bifidobacteria-mediated immune system imprinting early in life. *Cell* 2021;184(15):3884-
1581 3898.e11. doi: 10.1016/j.cell.2021.05.030.

1582 5. Bakos E, Homolya L. Portrait of multifaceted transporter, the multidrug resistance-
1583 associated protein 1 (MRP1/ABCC1). *Pflugers Arch* 2007;453(5):621-41. doi:
1584 10.1007/s00424-006-0160-8.

1585 6. He SM, Li R, Kanwar JR, Zhou SF. Structural and functional properties of human
1586 multidrug resistance protein 1 (MRP1/ABCC1). *Curr Med Chem* 2011;18(3):439-81. doi:
1587 10.2174/092986711794839197.

1588 7. Yoshii K, Hosomi K, Sawane K, Kunisawa J. Metabolism of Dietary and Microbial
1589 Vitamin B Family in the Regulation of Host Immunity. *Front Nutr* 2019;6:48. doi:
1590 10.3389/fnut.2019.00048.

1591 8. Beedholm-Ebsen R, van de Wetering K, Hardlei T, Nexø E, Borst P, Moestrup SK.
1592 Identification of multidrug resistance protein 1 (MRP1/ABCC1) as a molecular gate for
1593 cellular export of cobalamin. *Blood* 2010;115(8):1632-9. doi: 10.1182/blood-2009-07-
1594 232587.

1595 9. Lee JH, O'Sullivan DJ. Genomic insights into bifidobacteria. *Microbiol Mol Biol Rev*
1596 2010;74(3):378-416. doi: 10.1128/MMBR.00004-10.

1597 10. Bantel H, Schmitz ML, Raible A, Gregor M, Schulze-Osthoff K. Critical role of NF-
1598 kappaB and stress-activated protein kinases in steroid unresponsiveness. *FASEB J*
1599 2002;16(13):1832-4. doi: 10.1096/fj.02-0223fje.

1600 11. Karin M, Chang L. AP-1--glucocorticoid receptor crosstalk taken to a higher level. *J*
1601 *Endocrinol* 2001;169(3):447-51. doi: 10.1677/joe.0.1690447.

1602 12. Barbian ME, Owens JA, Naudin CR, Denning PW, Patel RM, Jones RM. Butyrate
1603 supplementation to pregnant mice elicits cytoprotection against colonic injury in the
1604 offspring. *Pediatr Res* 2021; doi: 10.1038/s41390-021-01767-1.

1605 13. Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, Hui KY, et al. Host-
1606 microbe interactions have shaped the genetic architecture of inflammatory bowel
1607 disease. *Nature* 2012;491(7422):119-24. doi: 10.1038/nature11582.

1608 14. Liu JZ, van Sommeren S, Huang H, Ng SC, Alberts R, Takahashi A, et al.
1609 Association analyses identify 38 susceptibility loci for inflammatory bowel disease and
1610 highlight shared genetic risk across populations. *Nat Genet* 2015;47(9):979-986. doi:
1611 10.1038/ng.3359.

1612 15. Ye BD, McGovern DPB. Genetic variation in IBD: progress, clues to pathogenesis
1613 and possible clinical utility. *Exp Rev Clin Immunol* 2016;12(10):1091-107. doi:
1614 10.1080/1744666X.2016.1184972.

1615 16. Di Narzo AF, Peters LA, Argmann C, Stojmirovic A, Perrigoue J, Li K, et al. Blood
1616 and Intestine eQTLs from an Anti-TNF-Resistant Crohn's Disease Cohort Inform IBD
1617 Genetic Association Loci. *Clin Transl Gastroenterol* 2016;7(6):e177. doi:
1618 10.1038/ctg.2016.34.

1619 17. Povoleri GAM, Nova-Lamperti E, Scottà C, Fanelli G, Chen YC, Becker PD, et al.
1620 Human retinoic acid-regulated CD161+ regulatory T cells support wound repair in
1621 intestinal mucosa. *Nat Immunol* 2018;19(12):1403-1414. doi: 10.1038/s41590-018-
1622 0230-z.

1623 18. Wang ZL, Wang YD, Wang K, Li JA, Li L. KFL2 participates in the development of
1624 ulcerative colitis through inhibiting inflammation via regulating cytokines. *Eur Rev Med*
1625 *Pharmacol Sci* 2018;2(15):4941-4948. doi: 10.26355/eurrev_201808_15633.

1626 19. Banerjee SS, Feinberg MW, Watanabe M, Gray S, Haspel RL, Denkinger DJ, et al.
1627 The Krüppel-like factor KLF2 inhibits peroxisome proliferator-activated receptor-gamma
1628 expression and adipogenesis. *J Biol Chem* 2003;278(4):2581-4. doi:
1629 10.1074/jbc.M210859200.

1630 20. Banerjee SS, Lin Z, Atkins GB, Greif DM, Rao RM, Kumar A, et al. KLF2 Is a novel
1631 transcriptional regulator of endothelial proinflammatory activation. *J Exp Med*
1632 2004;199(10):1305-15. doi: 10.1084/jem.20031132.

1633 21. Pabbisetty SK, Rabacal W, Volanakis EJ, Parekh VV, Olivares-Villagómez D,
1634 Cendron D, et al. Peripheral tolerance can be modified by altering KLF2-regulated Treg
1635 migration. *Proc Natl Acad Sci U S A* 2016;113(32):E4662-70. doi:
1636 10.1073/pnas.1605849113.

1637 22. Szigeti R, Pangas SA, Nagy-Szakal D, Dowd SE, Shulman RJ, Olive AP, et al.
1638 SMAD4 haploinsufficiency associates with augmented colonic inflammation in select
1639 humans and mice. *Ann Clin Lab Sci* 2012;42(4):401-8.

1640 23. Means AL, Freeman TJ, Zhu J, Woodbury LG, Marincola-Smith P, Wu C, et al.
1641 Epithelial Smad4 Deletion Up-Regulates Inflammation and Promotes Inflammation-
1642 Associated Cancer. *Cell Mol Gastroenterol Hepatol* 2018;6(3):257-276. doi:
1643 10.1016/j.jcmgh.2018.05.006.

1644 24. Wang Y, Xue Q, Zheng Q, Jin Y, Shen X, Yang M, et al. SMAD4 mutation correlates
1645 with poor prognosis in non-small cell lung cancer. *Lab Invest* 2021;101(4):463-476. doi:
1646 10.1038/s41374-020-00517-x.

1647 25. Mizuno T, Cloyd JM, Vicente D, Omichi K, Chun YS, Kopetz SE, et al. SMAD4 gene
1648 mutation predicts poor prognosis in patients undergoing resection for colorectal liver
1649 metastases. *Eur J Surg Oncol* 2018;44(5):684-692. doi: 10.1016/j.ejso.2018.02.247.

1650 26. Miyaki M, Iijima T, Konishi M, Sakai K, Ishii A, Yasuno M, et al. Higher frequency of
1651 Smad4 gene mutation in human colorectal cancer with distant metastasis. *Oncogene*
1652 1999;18(20):3098-103. doi: 10.1038/sj.onc.1202642.

1653 27. Lin LH, Chang KW, Cheng HW, Liu CJ. SMAD4 Somatic Mutations in Head and
1654 Neck Carcinoma Are Associated With Tumor Progression. *Front Oncol* 2019;9:1379.
1655 doi: 10.3389/fonc.2019.01379.

1656 28. Gart E, van Duyvenvoorde W, Toet K, Caspers MPM, Verschuren L, Nielsen MJ, et
1657 al. Butyrate Protects against Diet-Induced NASH and Liver Fibrosis and Suppresses
1658 Specific Non-Canonical TGF- β Signaling Pathways in Human Hepatic Stellate Cells.
1659 *Biomedicines* 2021;9(12):1954. doi: 10.3390/biomedicines9121954.

1660 29. Dong S, Tian Q, Zhu T, Wang K, Lei G, Liu Y, et al. SLC39A5 dysfunction impairs
1661 extracellular matrix synthesis in high myopia pathogenesis. *J Cell Mol Med*
1662 2021;25(17):8432-8441. doi: 10.1111/jcmm.16803.

1663 30. Ellinghaus D, Ellinghaus E, Nair RP, Stuart PE, Esko T, Metspalu A, et al.
1664 Combined analysis of genome-wide association studies for Crohn disease and psoriasis
1665 identifies seven shared susceptibility loci. *Am J Human Genet* 2012;90(4):636-47. doi:
1666 10.1016/j.ajhg.2012.02.020.

1667 31. Ye BD, Choi H, Hong M, Yun WJ, Low HQ, Haritunians T, et al. Identification of Ten
1668 Additional Susceptibility Loci for Ulcerative Colitis Through Immunochip Analysis in
1669 Koreans. *Inflamm Bowel Dis* 2016;22(1):13-9. doi: 10.1097/MIB.0000000000000584.

1670 32. Parmar AS, Lappalainen M, Paavola-Sakki P, Halme L, Färkkilä M, Turunen U, et
1671 al. Association of celiac disease genes with inflammatory bowel disease in Finnish and
1672 Swedish patients. *Genes Immun* 2012;13(6):474-80. doi: 10.1038/gene.2012.21.

1673 33. Verma SC, Mahadevan S. The chbG gene of the chitobiose (chb) operon of
1674 *Escherichia coli* encodes a chitooligosaccharide deacetylase. *J Bacteriol*
1675 2012;194(18):4959-71. doi: 10.1128/JB.00533-12.

1676 34. Kim EJ, Park MK, Kang GJ, Byun HJ, Kim HJ, Yu L, et al. YDJC Induces Epithelial-
1677 Mesenchymal Transition via Escaping from Interaction with CDC16 through
1678 Ubiquitination of PP2A. *J Oncol* 2019;2019:3542537. doi: 10.1155/2019/3542537.

1679 35. Kim EJ, Park MK, Byun HJ, Kang GJ, Yu L, Kim HJ, et al. YdjC chitooligosaccharide
1680 deacetylase homolog induces keratin reorganization in lung cancer cells: involvement of

1681 interactions between YDJC and CDC16. *Oncotarget* 2018;9(33):22915-22928. doi:
1682 10.18632/oncotarget.25145.

1683 36. Parker BJ, Wearsch PA, Veloo ACM, Rodriguez-Palacios A. The Genus Alistipes:
1684 Gut Bacteria With Emerging Implications to Inflammation, Cancer, and Mental Health.
1685 *Front Immunol* 2020;11:906. doi: 10.3389/fimmu.2020.00906.

1686 37. Dziarski R, Park SY, Kashyap DR, Dowd SE, Gupta D. Pglyrp-Regulated Gut
1687 Microflora *Prevotella falsenii*, *Parabacteroides distasonis* and *Bacteroides eggerthii*
1688 Enhance and Alistipes *finegoldii* Attenuates Colitis in Mice. *PLoS One*
1689 2016;11(1):e0146162. doi: 10.1371/journal.pone.0146162.

1690 38. Butera A, Di Paola M, Pavarini L, Strati F, Pindo M, Sanchez M, et al. Nod2
1691 Deficiency in mice is Associated with Microbiota Variation Favouring the Expansion of
1692 mucosal CD4⁺ LAP⁺ Regulatory Cells. *Sci Rep* 2018;8(1):14241. doi: 10.1038/s41598-
1693 018-32583-z.

1694 39. Li J, Sung CY, Lee N, Ni Y, Pihlajamäki J, Panagiotou G, et al. Probiotics modulated
1695 gut microbiota suppresses hepatocellular carcinoma growth in mice. *Proc Natl Acad Sci
1696 U S A* 2016;113(9):E1306-15. doi: 10.1073/pnas.1518189113.

1697 40. Rodriguez-Palacios A, Kodani T, Kaydo L, Pietropaoli D, Corridoni D, Howell S, et
1698 al. Stereomicroscopic 3D-pattern profiling of murine and human intestinal inflammation
1699 reveals unique structural phenotypes. *Nat Commun* 2015;6:7577. doi:
1700 10.1038/ncomms8577.

1701 41. Rodriguez-Palacios A, Harding A, Menghini P, Himmelman C, Retuerto M,
1702 Nickerson KP, et al. The Artificial Sweetener Splenda Promotes Gut Proteobacteria,
1703 Dysbiosis, and Myeloperoxidase Reactivity in Crohn's Disease-Like Ileitis. *Inflamm
1704 Bowel Dis* 2018;24(5):1005-1020. doi: 10.1093/ibd/izy060.

1705 42. Grondin JA, Kwon YH, Far PM, Haq S, Khan WI. Mucins in Intestinal Mucosal
1706 Defense and Inflammation: Learning From Clinical and Experimental Studies. *Front
1707 Immunol* 2020;11:2054. doi: 10.3389/fimmu.2020.02054.

1708 43. Das S, Rachagani S, Sheinin Y, Smith LM, Gurumurthy CB, Roy HK, et al. Mice
1709 deficient in Muc4 are resistant to experimental colitis and colitis-associated colorectal
1710 cancer. *Oncogene* 2016;35(20):2645-54. doi: 10.1038/onc.2015.327.

1711 44. Lam YY, Ha CWY, Campbell CR, Mitchell AJ, Dinudom A, Oscarsson J, et al.
1712 Increased gut permeability and microbiota change associate with mesenteric fat
1713 accumulation and metabolic dysfunction in diet-induced obese mice. *PLoS One*
1714 2012;7(3):e34233. doi: 10.1371/journal.pone.0034233.

1715 45. Peng Y, Yan Y, Wan P, Chen D, Ding Y, Ran L, et al. Gut microbiota modulation
1716 and anti-inflammatory properties of anthocyanins from the fruit of *Lycium ruthenicum*
1717 Murray in dextran sodium sulfate-induced colitis in mice. *Free Radic Biol Med*
1718 2019;136:96-108. doi: 10.1016/j.freeradbiomed.2019.04.005.

1719 46. Golubeva AV, Crampton S, Desbonnet L, Edge D, O'Sullivan O, Lomasney KW, et
1720 al. Prenatal stress-induced alterations in major physiological systems correlate with gut
1721 microbiota composition in adulthood. *Psychoneuroendocrinology* 2015;60:58-74. doi:
1722 10.1016/j.psyneuen.2015.06.002.

1723 47. Metwaly A, Dunkel A, Waldschmitt N, Raj ACD, Lagkouvardos I, Corraliza AM, et al.
1724 Integrated microbiota and metabolite profiles link Crohn's disease to sulfur metabolism.
1725 *Nat Commun* 2020;11(1):4322. doi: 10.1038/s41467-020-17956-1.

1726 48. Ryan FJ, Ahern AM, Fitzgerald RS, Laserna-Mendieta EJ, Power EM, Clooney AG,
1727 et al. Colonic microbiota is associated with inflammation and host epigenomic
1728 alterations in inflammatory bowel disease. *Nat Commun* 2020;11(1):1512. doi:
1729 10.1038/s41467-020-15342-5.

1730 49. Chollangi S, Mather T, Rodgers KK, Ash JD. A unique loop structure in oncostatin M
1731 determines binding affinity toward oncostatin M receptor and leukemia inhibitory factor
1732 receptor. *J Biol Chem* 2012;287(39):32848-59. doi: 10.1074/jbc.M112.387324.

1733 50. Verstockt S, Verstockt B, Machiels K, Vancamelbeke M, Ferrante M, Cleynen I, et
1734 al. Oncostatin M Is a Biomarker of Diagnosis, Worse Disease Prognosis, and
1735 Therapeutic Nonresponse in Inflammatory Bowel Disease. *Inflamm Bowel Dis*
1736 2021;27(10):1564-1575. doi: 10.1093/ibd/izab032.

1737 51. West NR, Hegazy AN, Owens BMJ, Bullers SJ, Linggi B, Buonocore S, et al.
1738 Oncostatin M drives intestinal inflammation and predicts response to tumor necrosis
1739 factor-neutralizing therapy in patients with inflammatory bowel disease. *Nat Med*
1740 2017;23(5):579-589. doi: 10.1038/nm.4307.

1741 52. Place SP. Single-point mutation in a conserved TPR domain of Hip disrupts
1742 enhancement of glucocorticoid receptor signaling. *Cell Stress Chaperones*
1743 2011;16(4):469-74. doi: 10.1007/s12192-010-0254-2.

1744 53. Bai R, Shi Z, Zhang JW, Li D, Zhu YL, Zheng S. ST13, a proliferation regulator,
1745 inhibits growth and migration of colorectal cancer cell lines. *J Zhejiang Univ Sci B*
1746 2012;13(11):884-893. doi: 10.1631/jzus.B1200037.

1747 54. Wang LB, Zheng S, Zhang SZ, Peng JP, Ye F, Fang SC, et al. Expression of ST13
1748 in colorectal cancer and adjacent normal tissues. *World J Gastroenterol*
1749 2005;11(3):336-9. doi: 10.3748/wjg.v11.i3.336.

1750

1751 **Box 2. Miscellaneous component pairs from sparse-CCA analysis.**

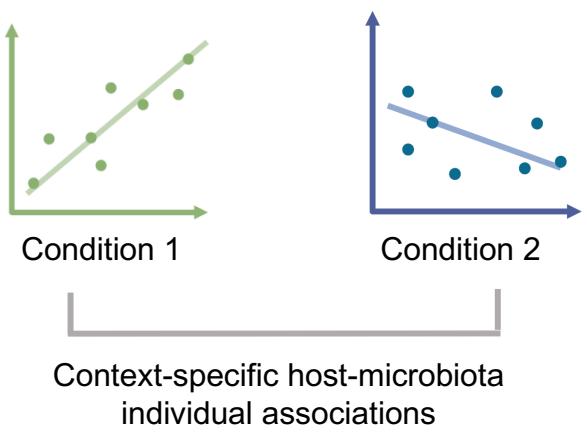
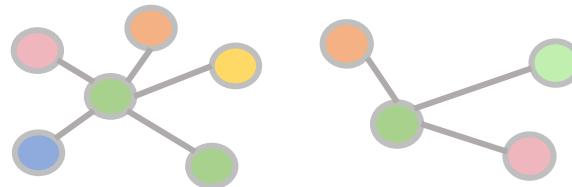
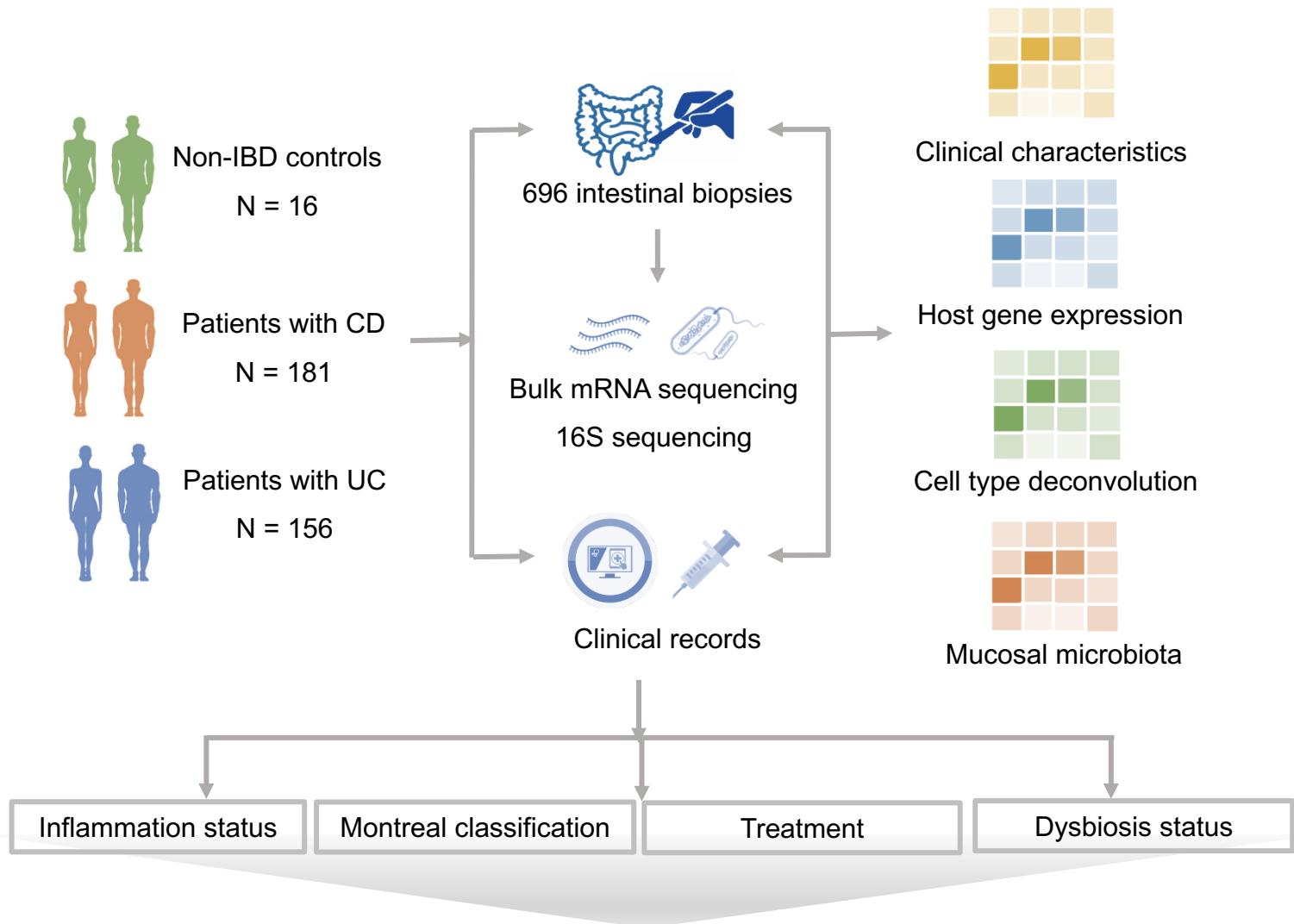
1752 The microbial part of the fifth pair of components (component pair 5, $P=2.87 \times 10^{-8}$,
1753 FDR<0.05) was formed by *Christensenellaceae*, *Ruminococcaceae*, *Lachnospiraceae*
1754 (NK4A136 group), Coriobacteria and the genera *Coprococcus* and *Ruminoclostridium*,
1755 which are all inversely associated with pathways representing SLC-mediated
1756 transmembrane transport (e.g. transport of bile acids and organic acids, metal ions and
1757 amine compounds) as well as biological oxidation and fat metabolism pathways
1758 including arachidonic acid metabolism and (glycero)phospholipid biosynthesis
1759 (**Supplementary Tables S15-S16**).

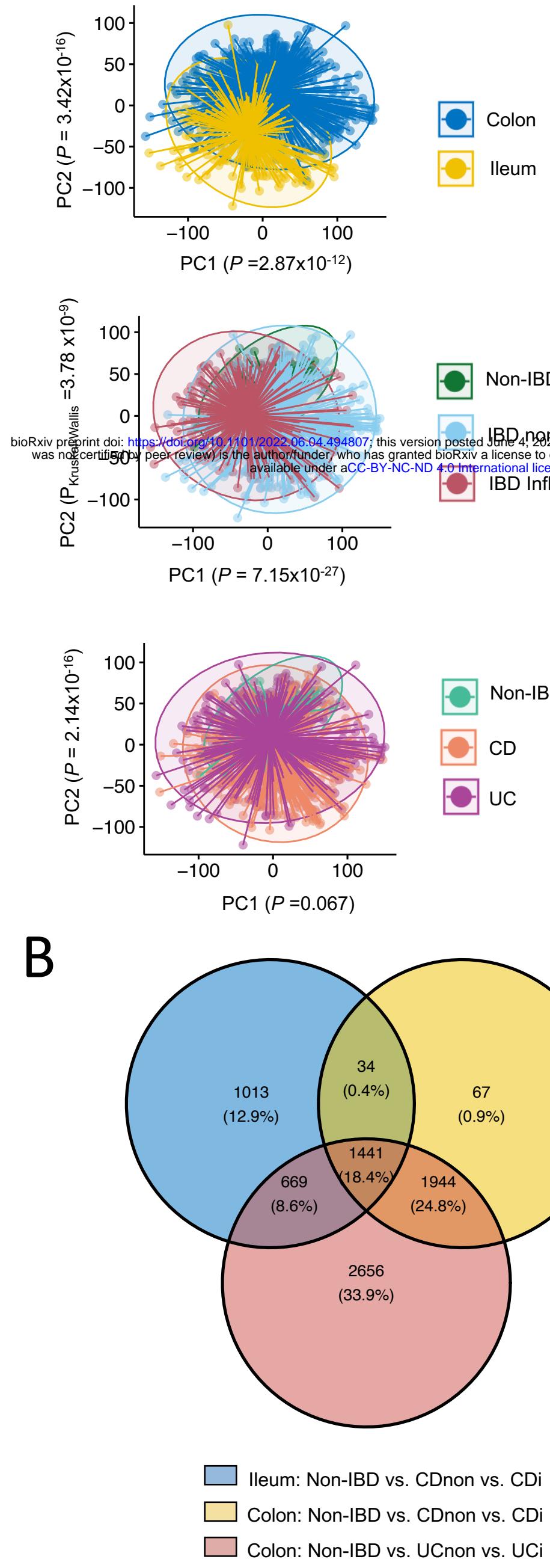
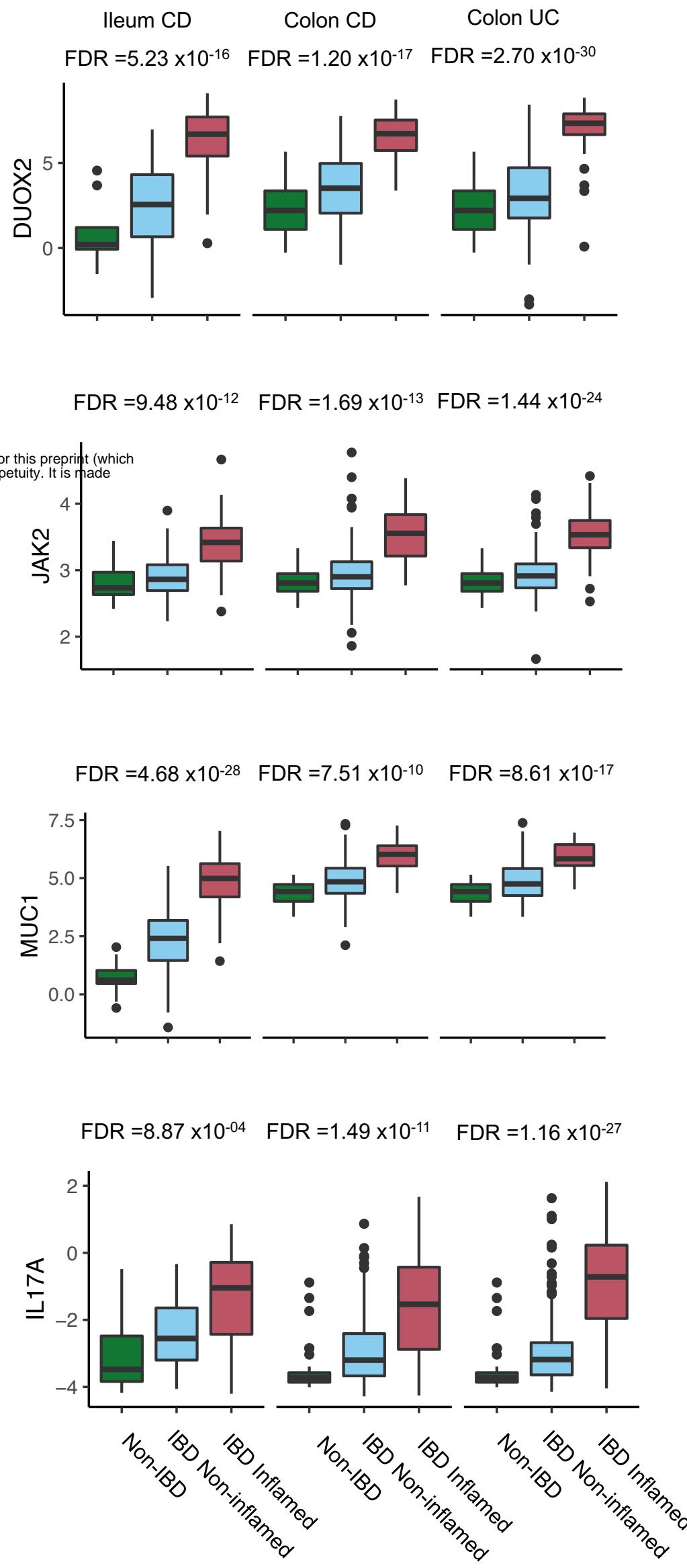
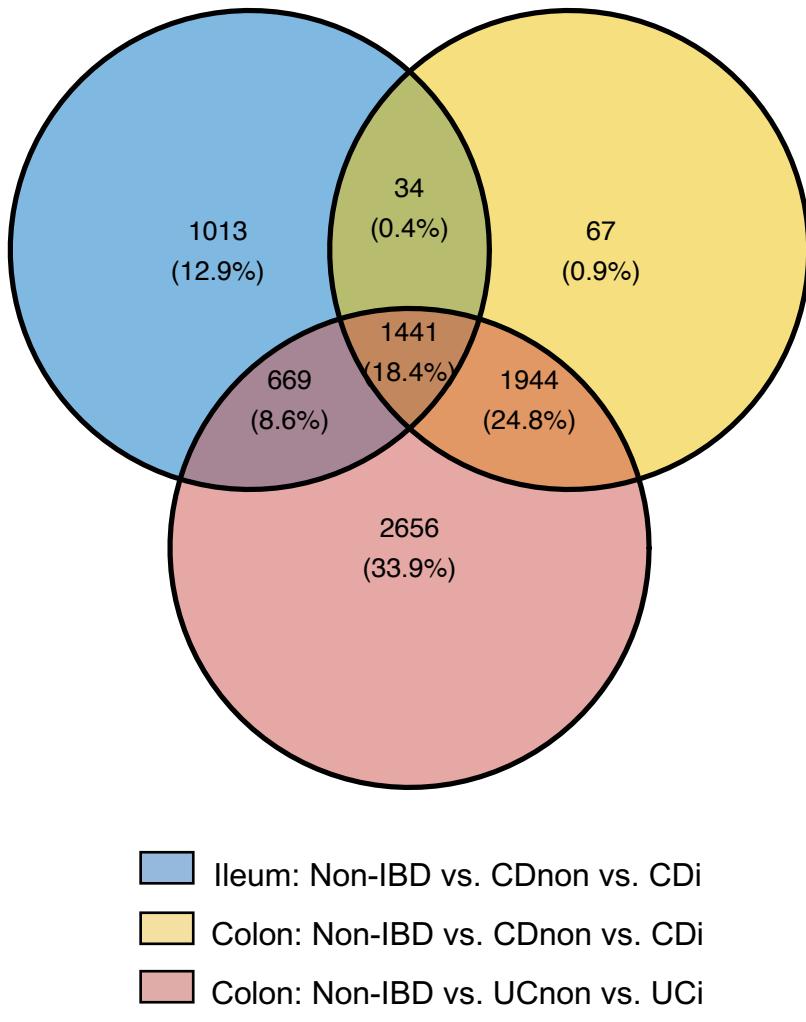
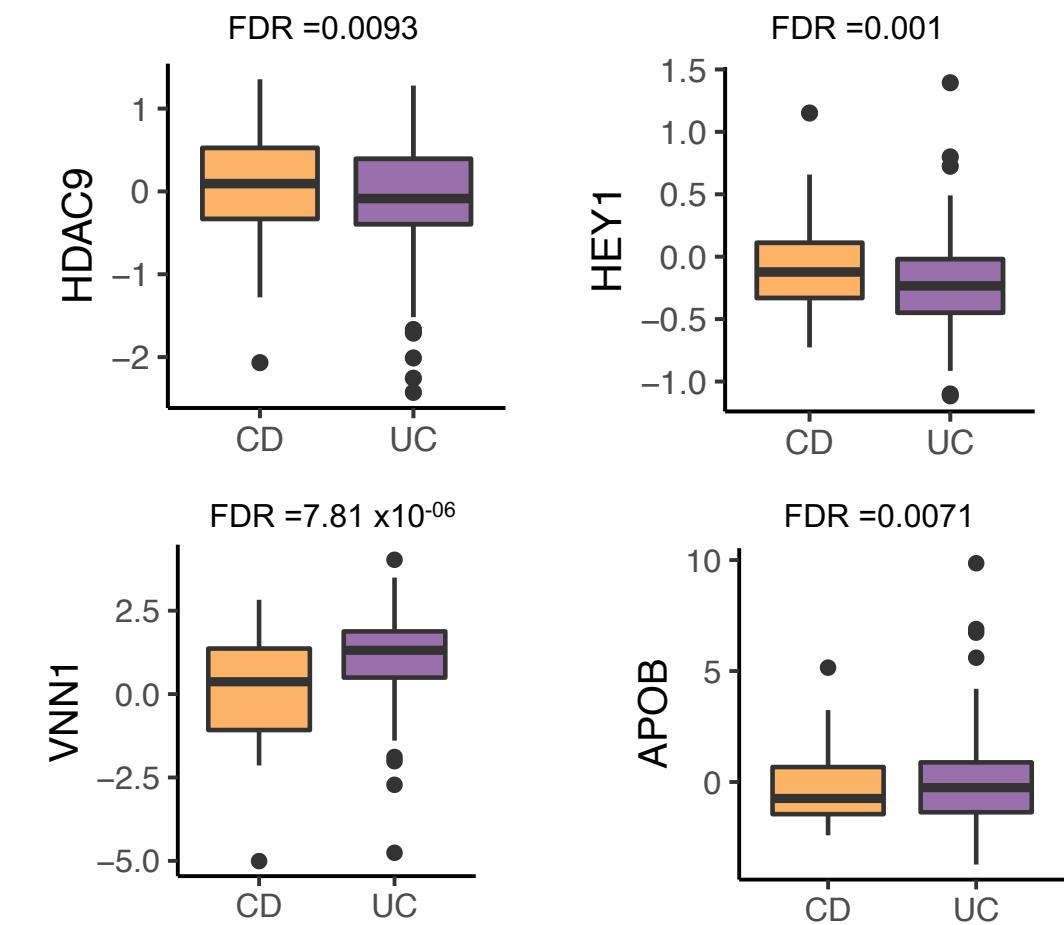
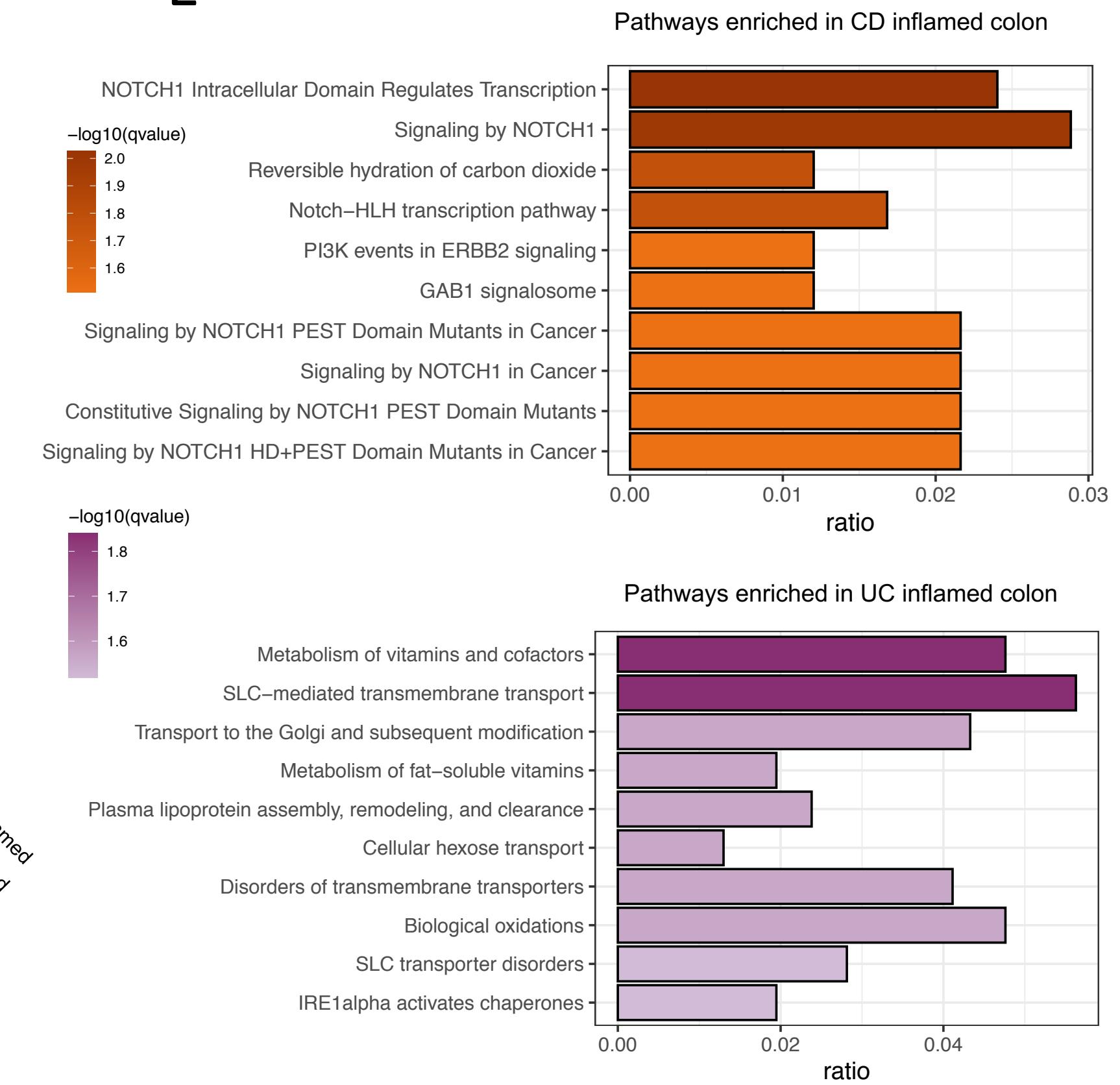
1760 In the sixth pair of components (component pair 9, $P=9.65 \times 10^{-7}$, FDR<0.05), the
1761 microbial component was primarily composed of bifidobacteria (i.e. order
1762 *Bifidobacteriales*, family *Bifidobacteriaceae* and genus *Bifidobacterium*), which were
1763 inversely associated with pathways representing phospholipid synthesis (e.g.
1764 phosphatidic acid synthesis) and NR1H2/NR1H3 or liver X receptor (LXR)-mediated

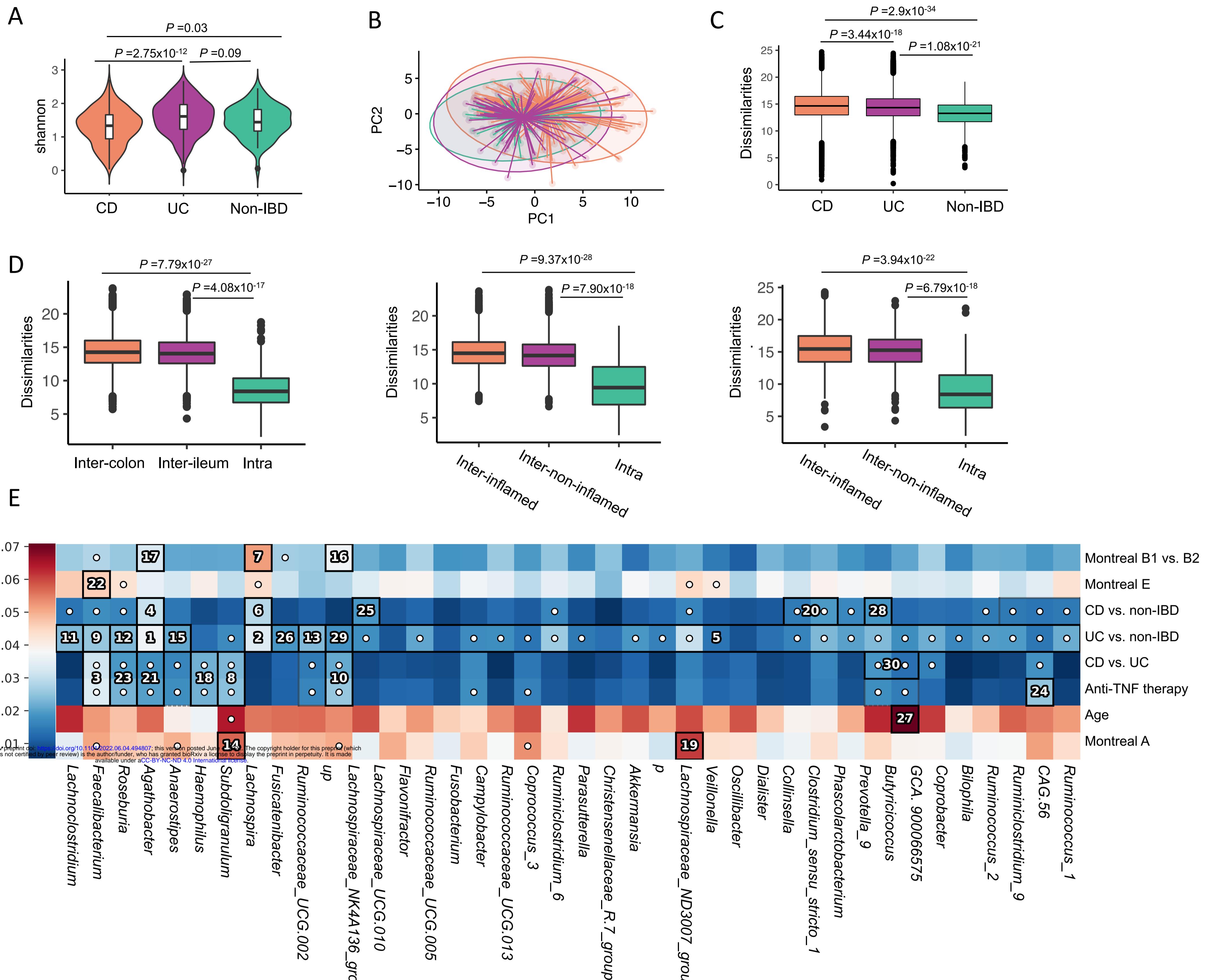
1765 signaling (**Supplementary Tables S17-S18**). NR1H3 (LXR- α) and NR1H2 (LXR- β) are
1766 ligand-activated transcription factors stimulated by endogenously produced oxysterols,
1767 which are in turn produced by oxidation of cholesterol, enzymatic reactions or
1768 alimentary processes [1]. Under physiological conditions, oxysterols are formed
1769 proportional to the cellular cholesterol content and thereby stimulate LXRs (acting as
1770 cholesterol sensors) to alter gene expression and activate protective mechanisms to
1771 prevent cholesterol overload in the cell. This occurs via inhibition of intestinal cholesterol
1772 absorption, activation of cholesterol efflux from cells to HDL (via ABCA1 and ABCG1
1773 transporters) and activation of the hepatic conversion of cholesterol to bile acids and
1774 stimulation of biliary cholesterol and bile acid excretion. In addition, LXR-agonists
1775 enhance *de novo* synthesis of fatty acids by stimulating the expression of the lipogenic
1776 transcription factor SREBP-1c, which may result in elevated plasma triglycerides and
1777 hepatic steatosis. LXRs are also involved in modulation of innate and adaptive immune
1778 responses and regulate diverse aspects of inflammatory gene expression in
1779 macrophages. The ability of LXRs to coordinate metabolic and immune response
1780 constitutes an attractive therapeutic target for treatment of IBD.

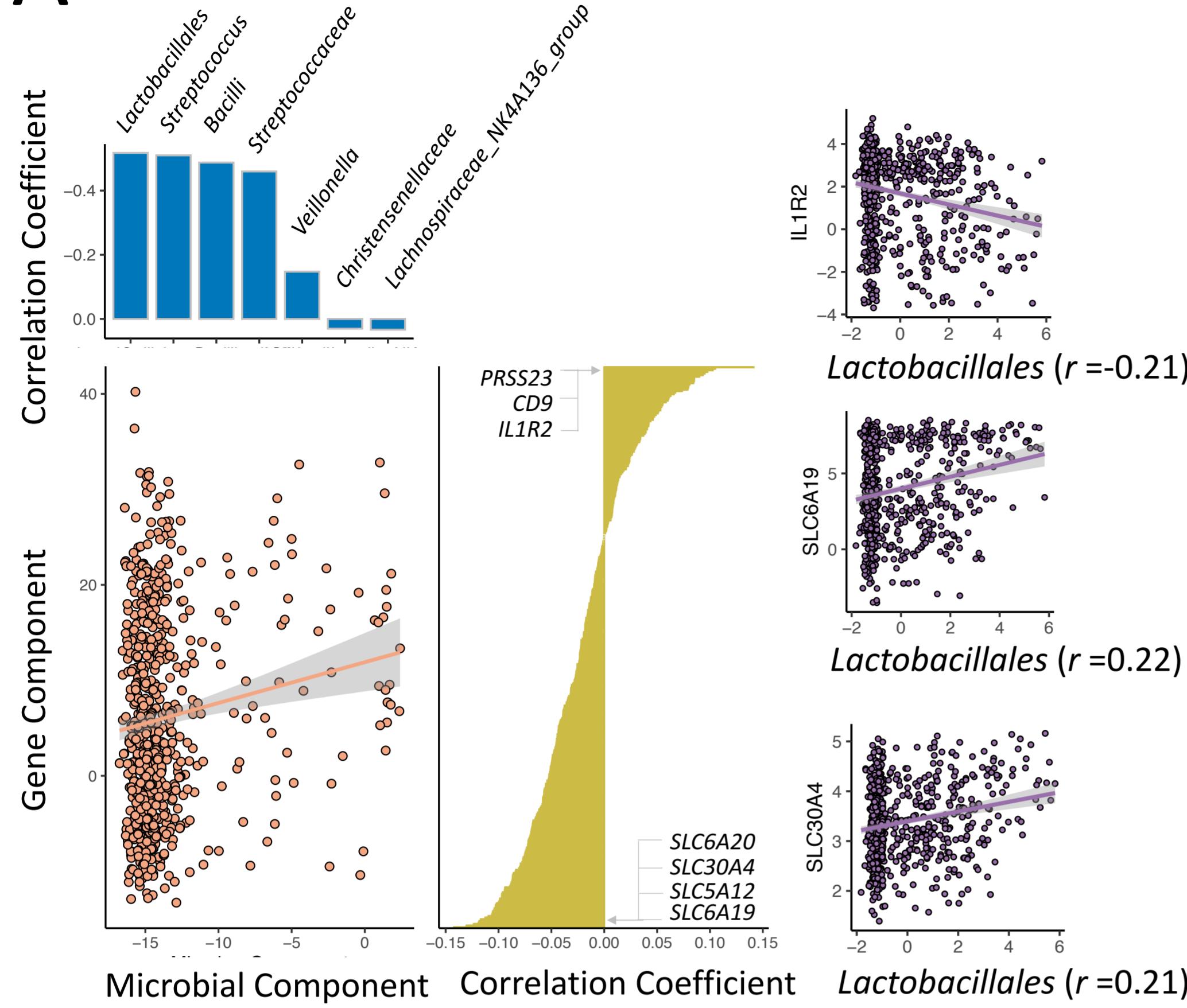
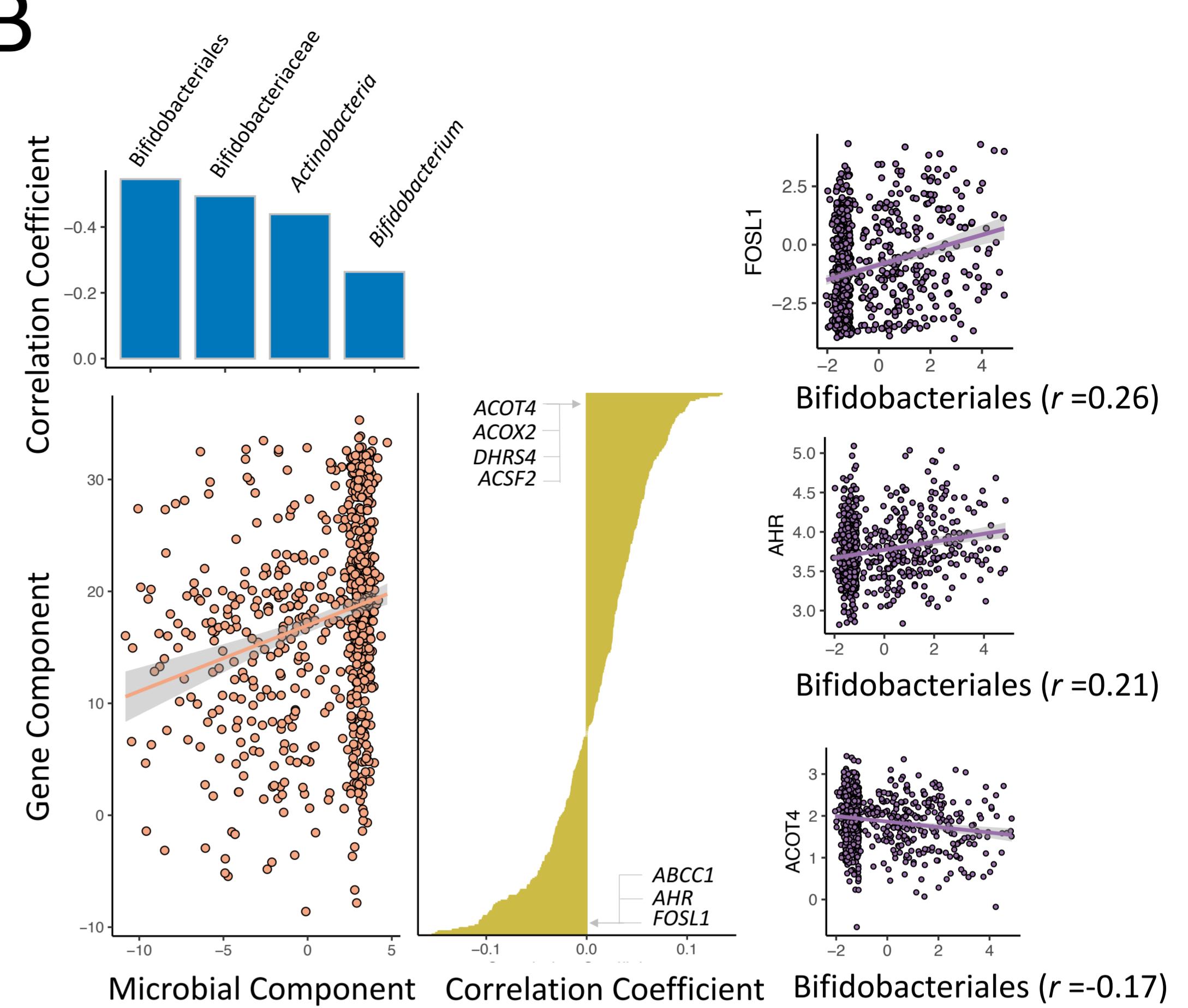
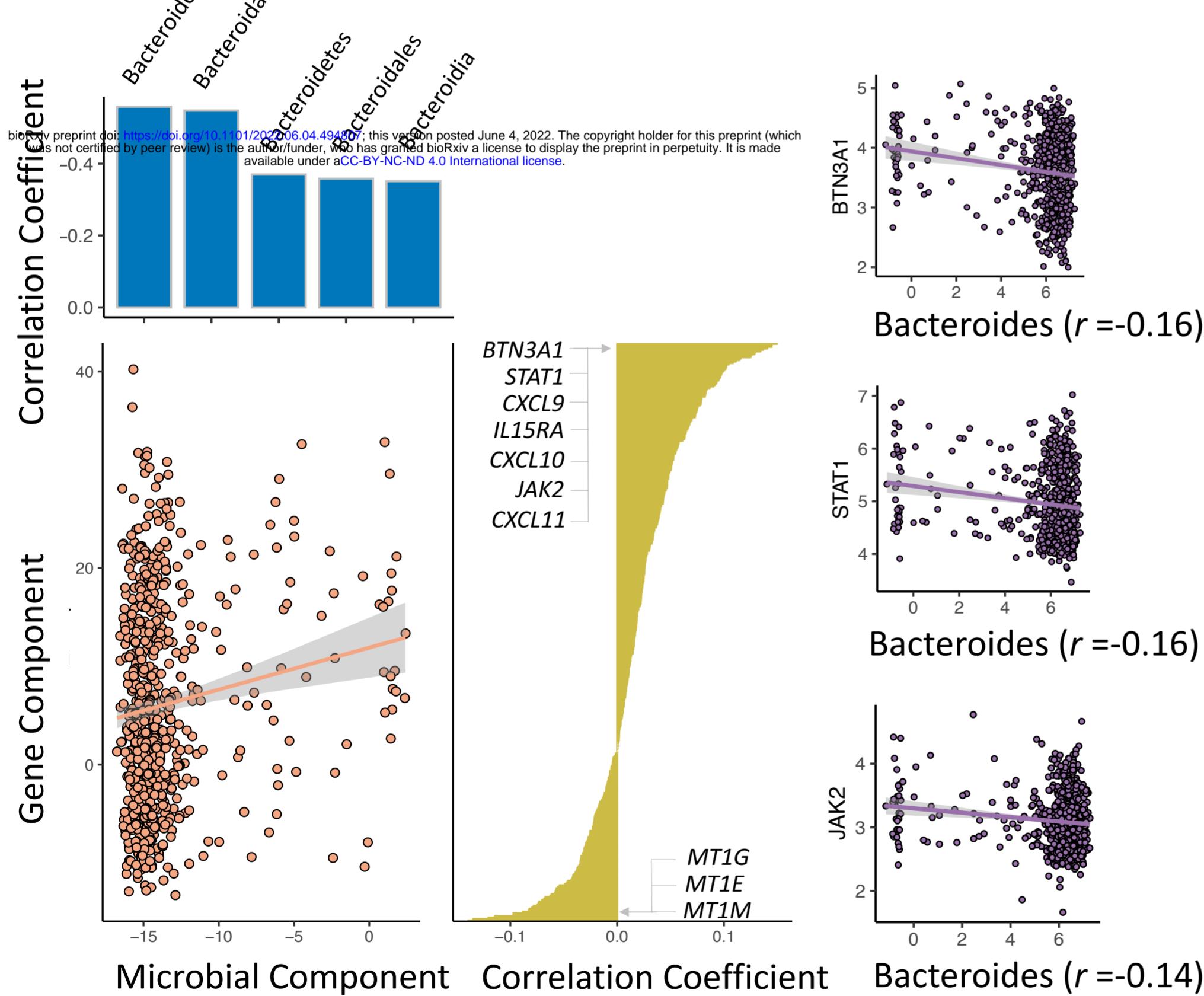
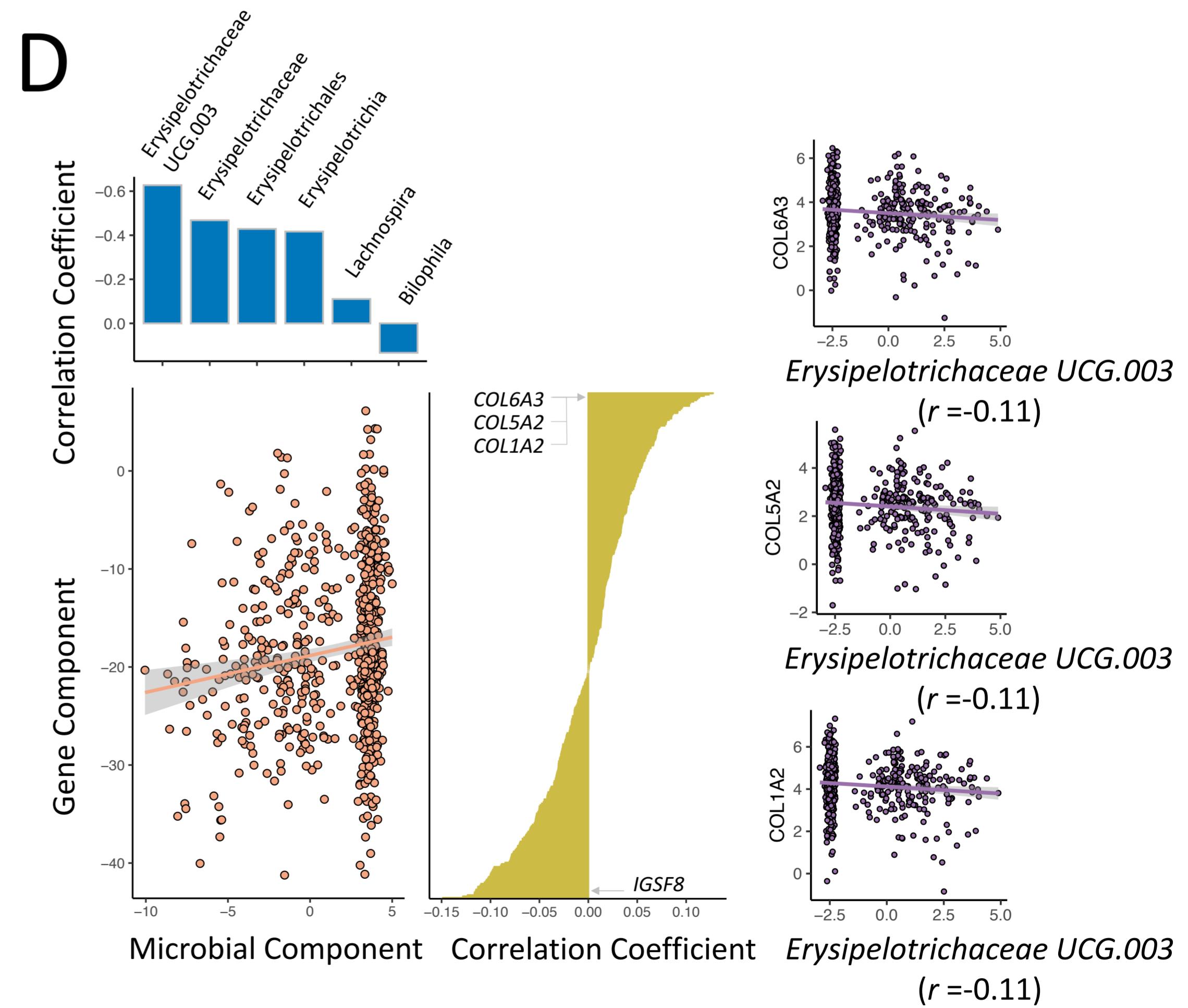
1781

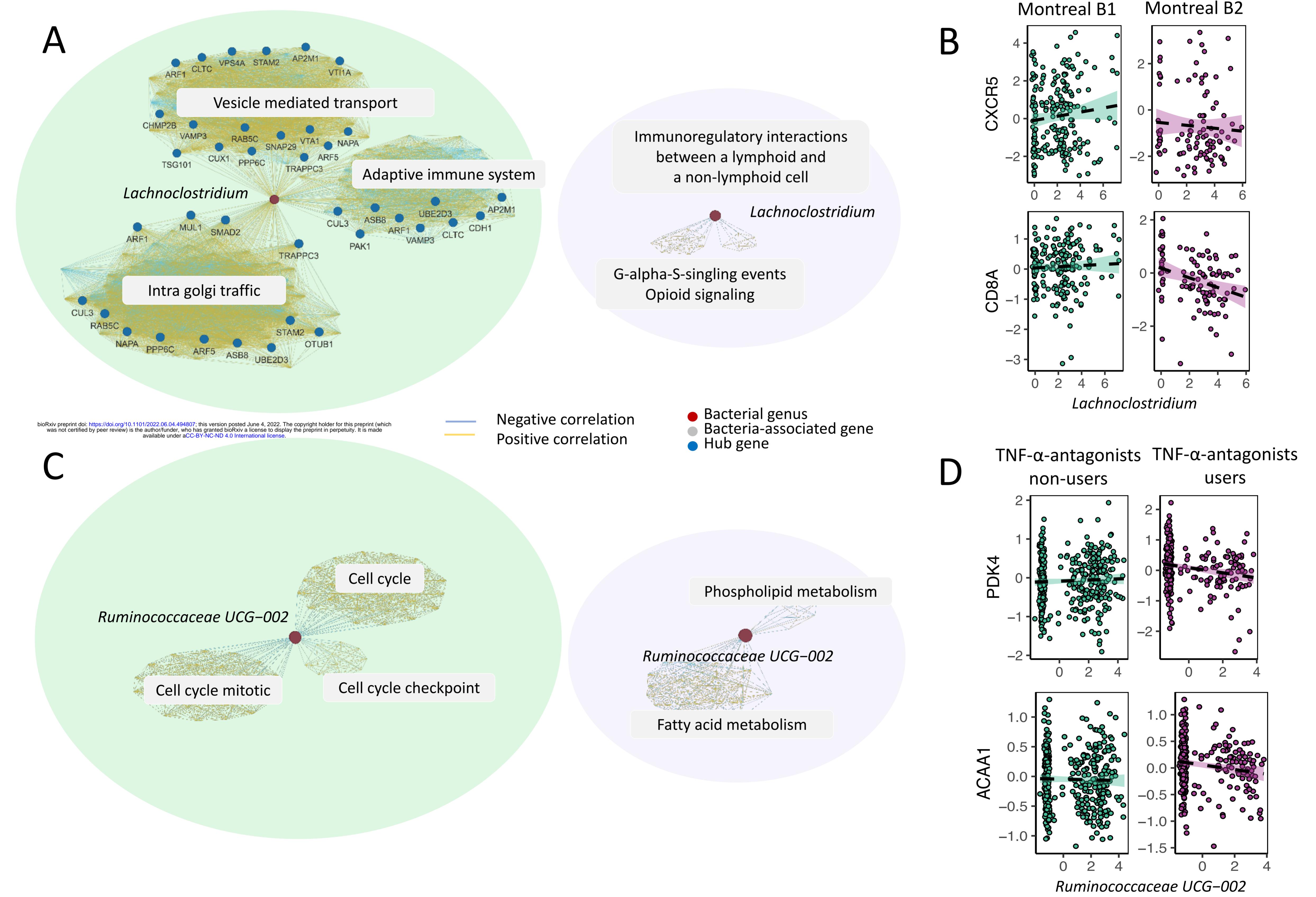
1782 1. Jakobsson T, Treuter E, Gustafsson JA, Steffensen KR. Liver X receptor biology and
1783 pharmacology: new pathways, challenges and opportunities. *Trends Pharmacol Sci*
1784 2012;33(7):394-404. doi: 10.1016/j.tips.2012.03.013.

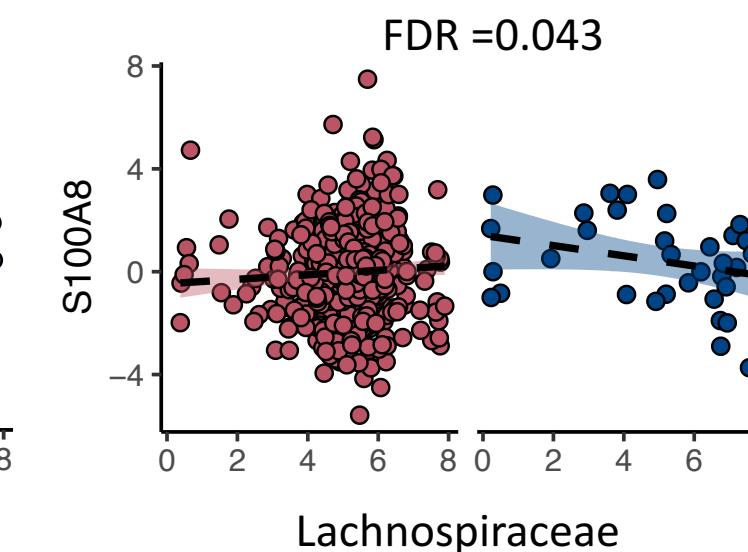
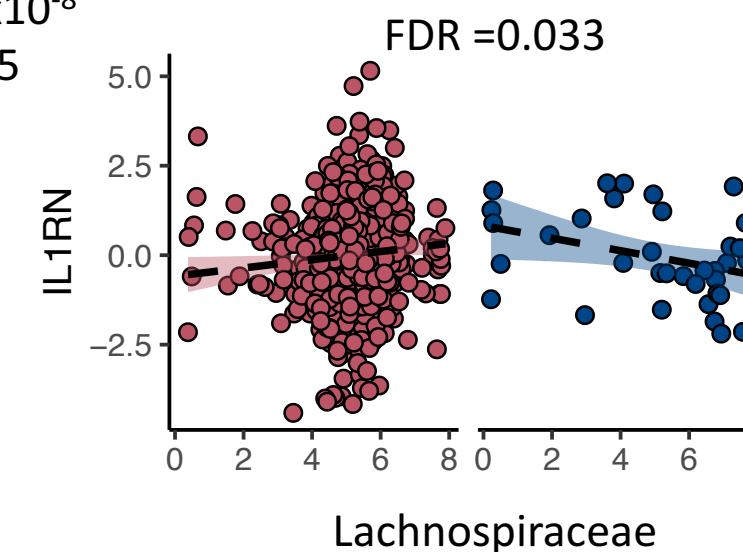
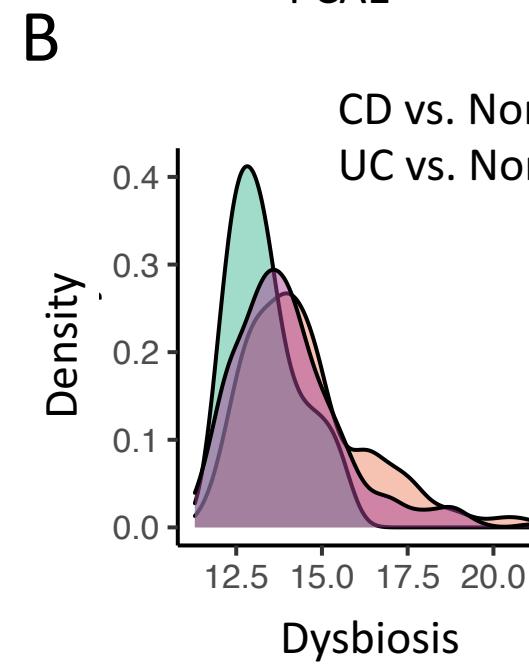
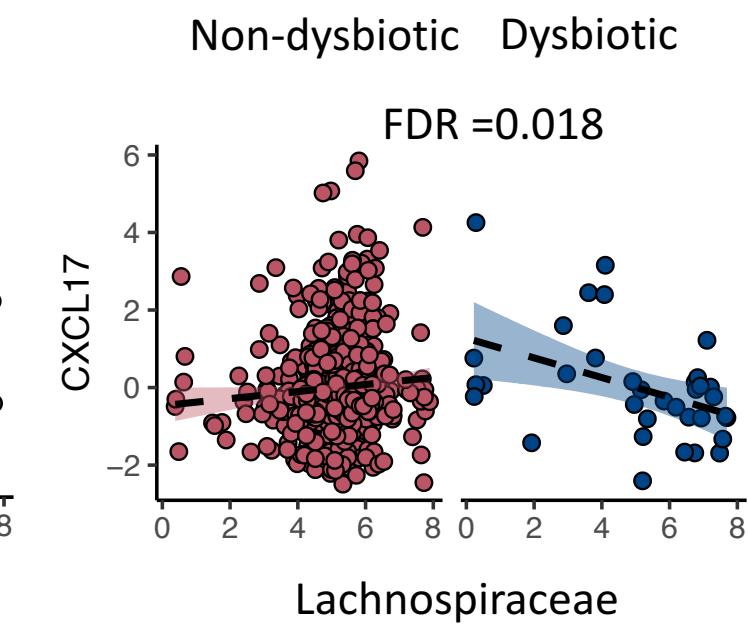
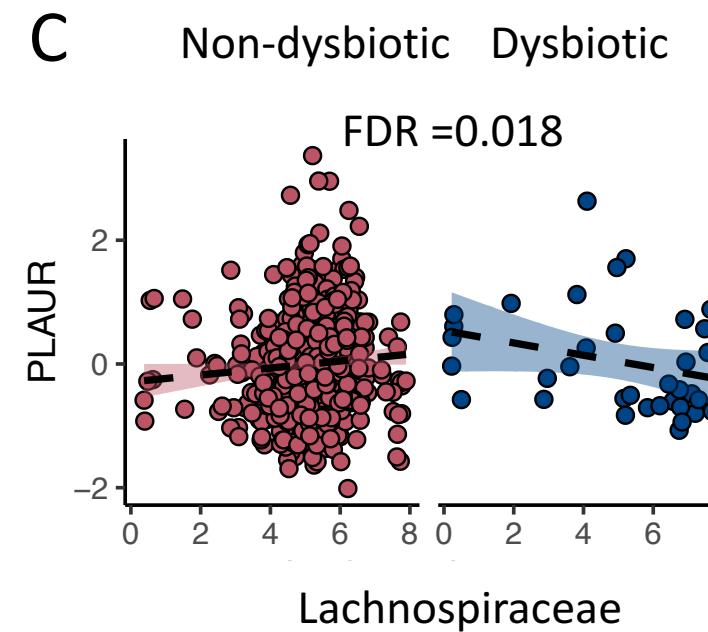
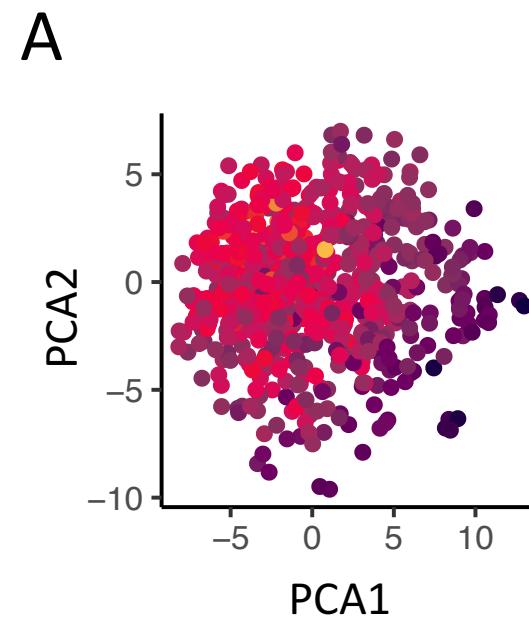


A**C****B****D****E**

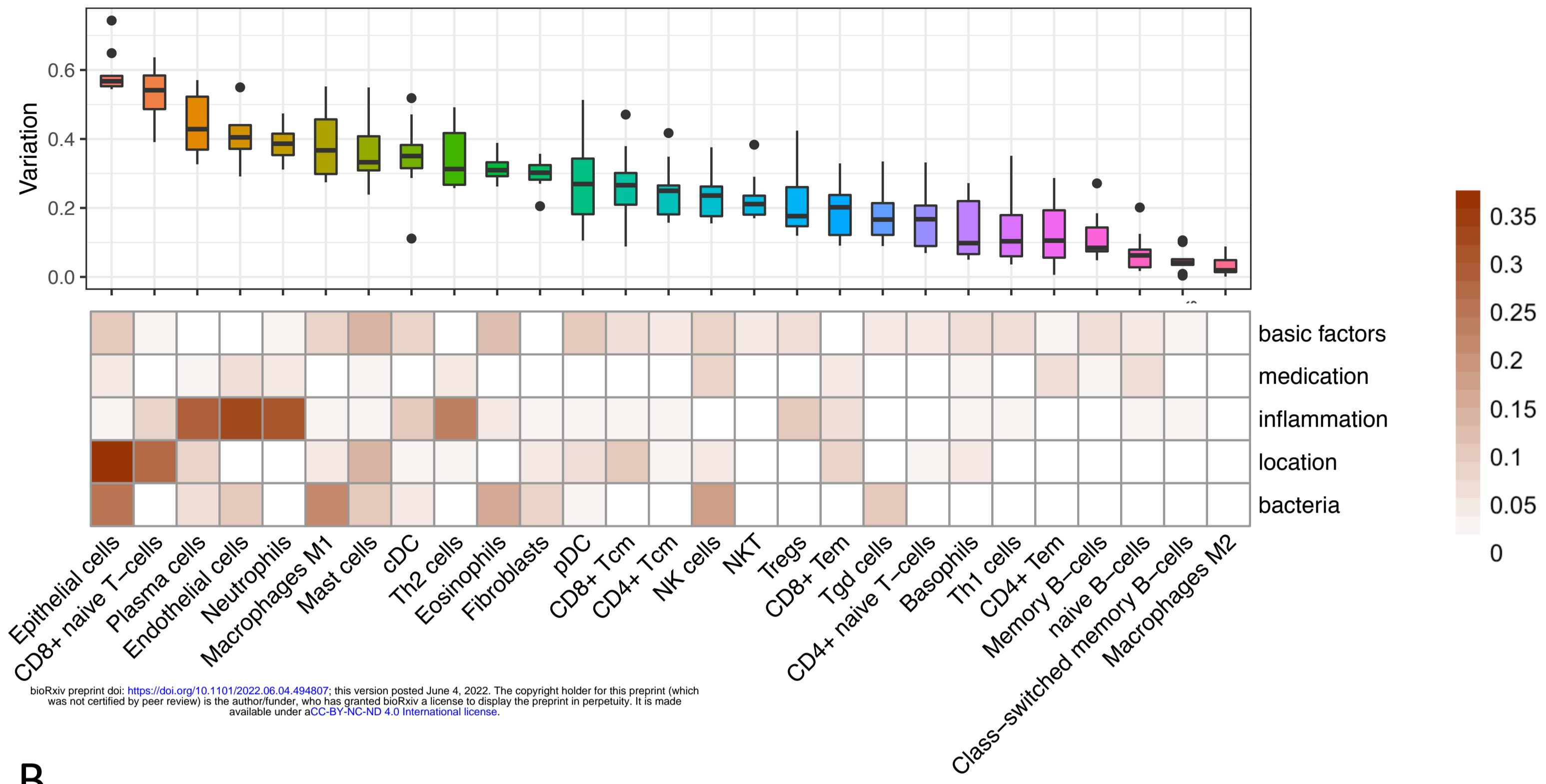


A**B****C****D**

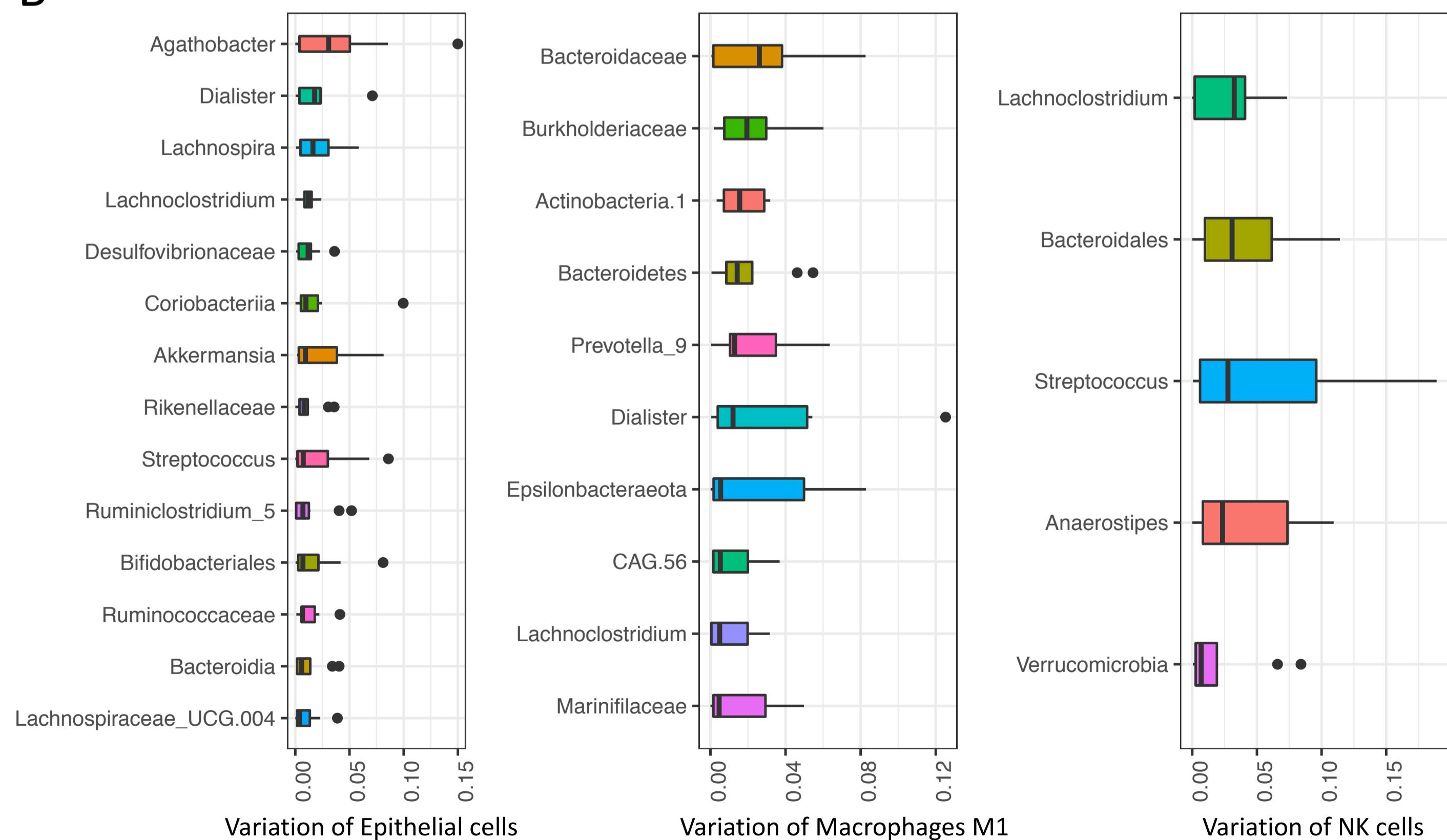




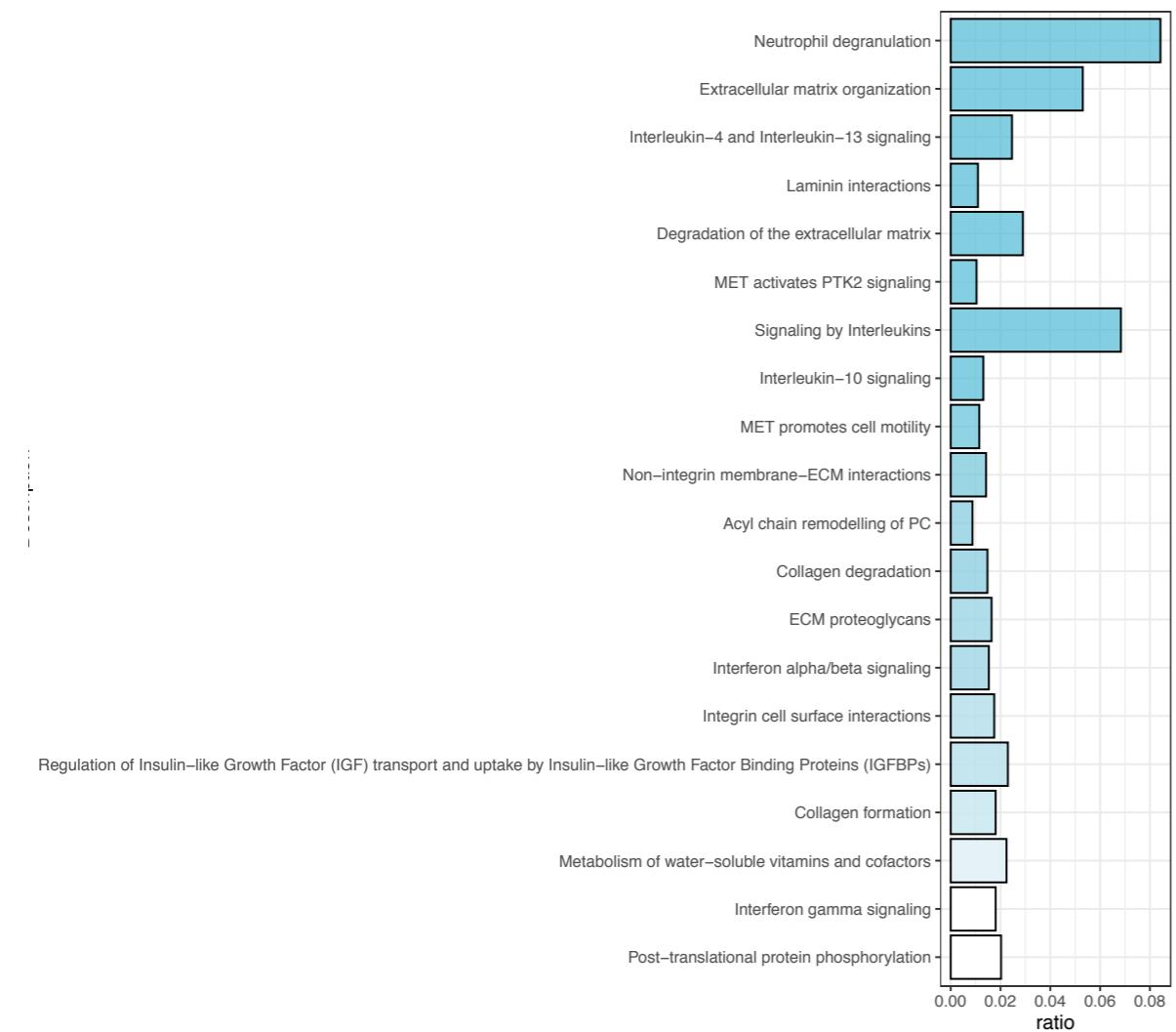
A



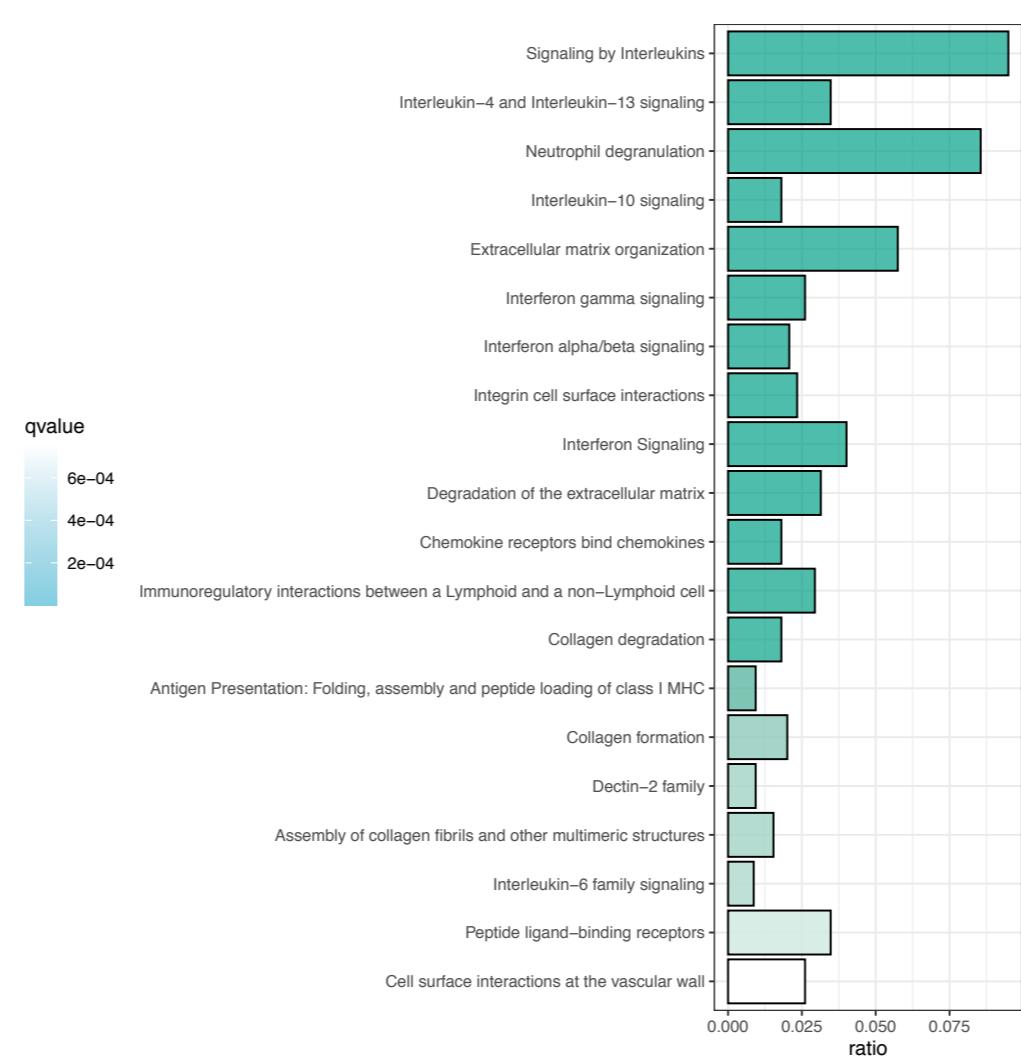
B



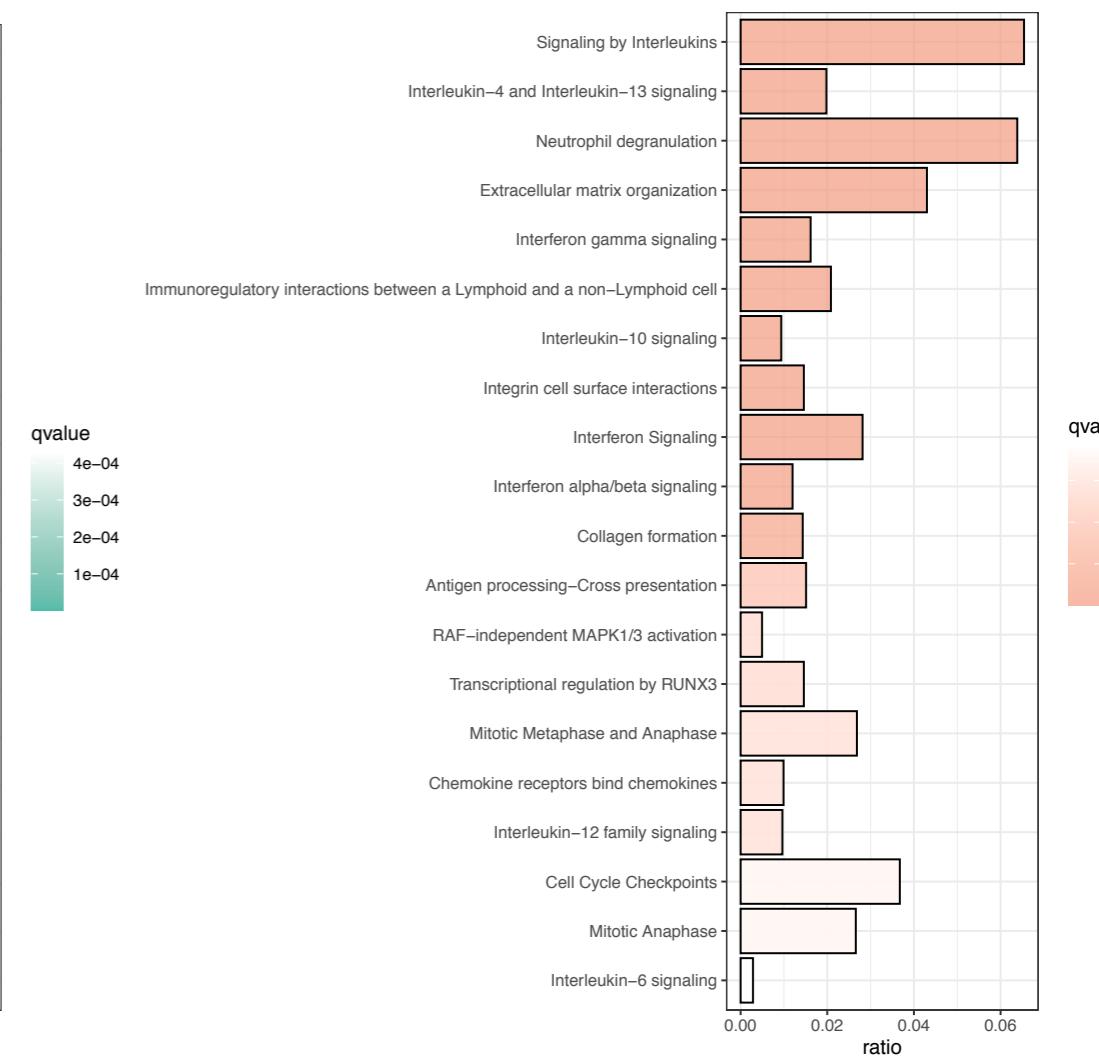
Ileum tissue: non-IBD vs. CDnon vs. CDi



Colon tissue: non-IBD vs. CDnon vs. CDi



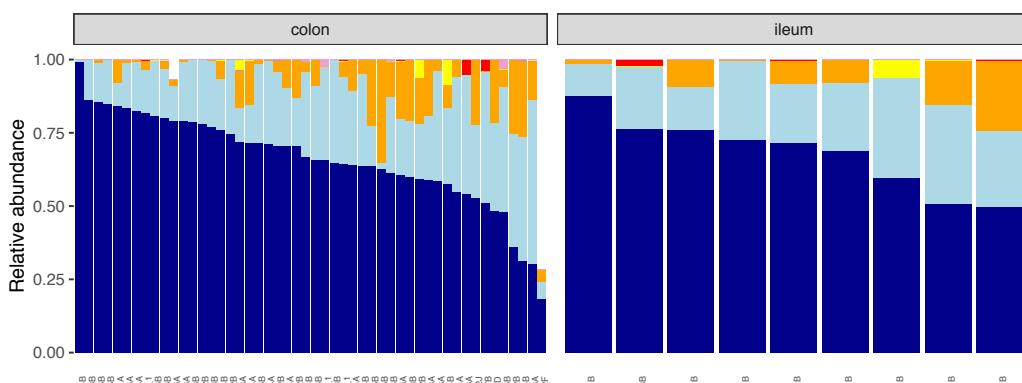
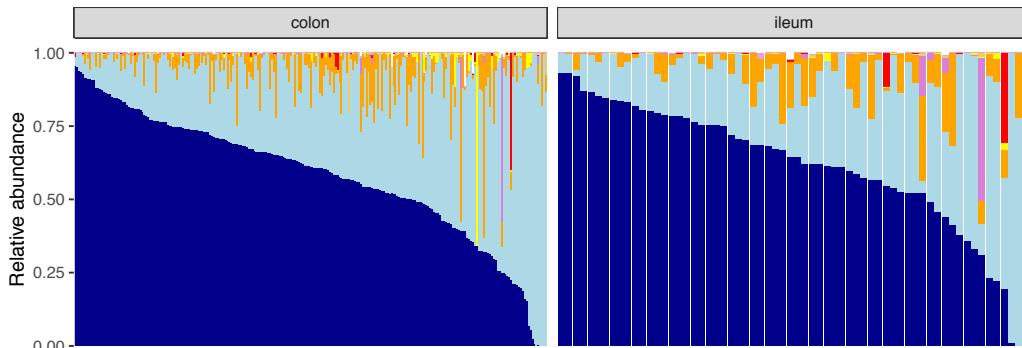
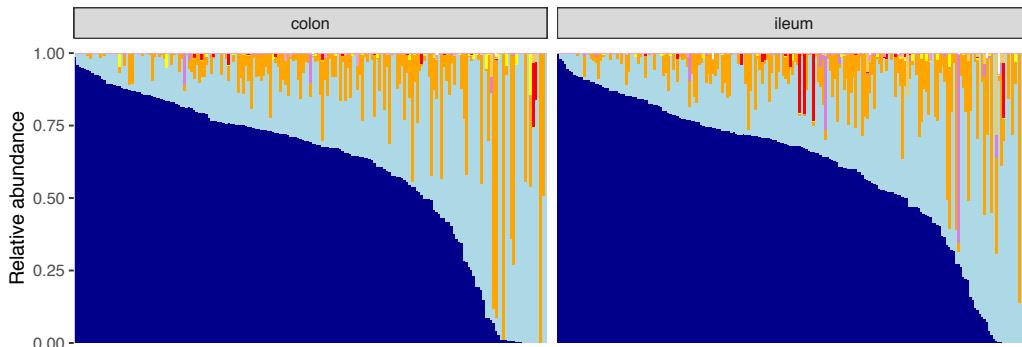
Colon tissue: non-IBD vs. UCnon vs. UCI



CD

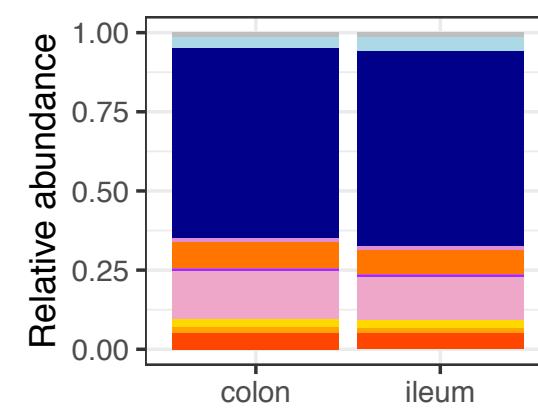
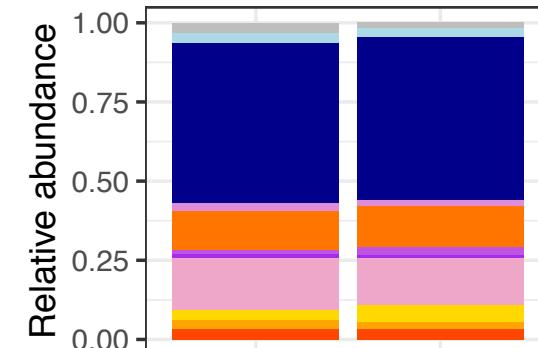
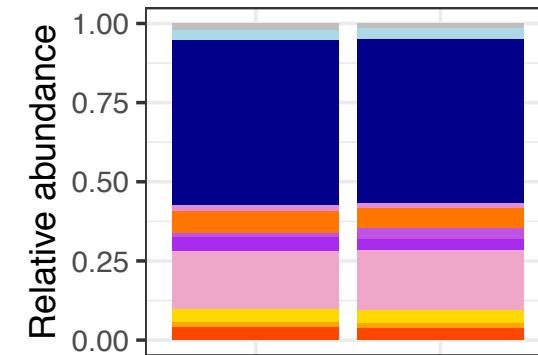
UC

Control

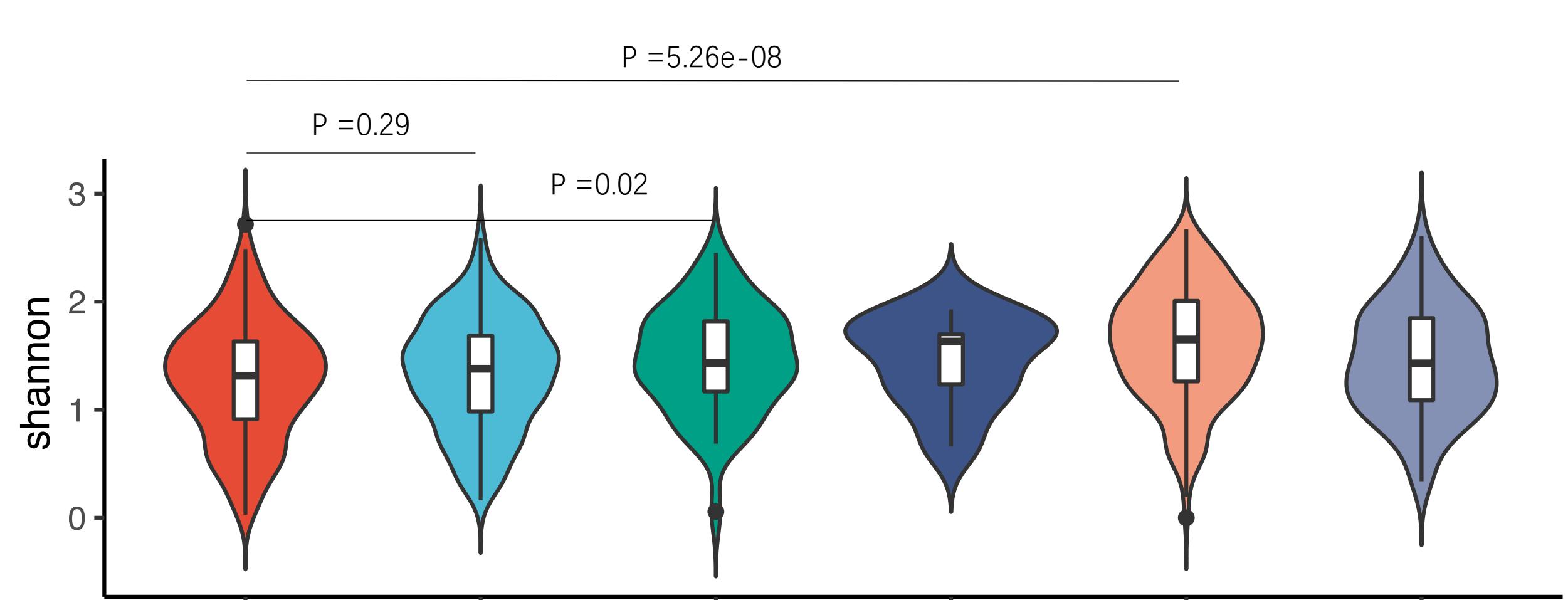


Level

Bacteroidetes	Proteobacteria	Actinobacteria	Epsilonbacteraeota
Firmicutes	Fusobacteria	Verrucomicrobia	Cyanobacteria



Bacteroides
Faecalibacterium
Sutterella
Parabacteroides
Alistipes
Lachnoclostridium
Agathobacter
Roseburia
Blautia
Haemophilus
Others



CD colon

CD ileum

Non-IBD
colon

Non-IBD
ileum

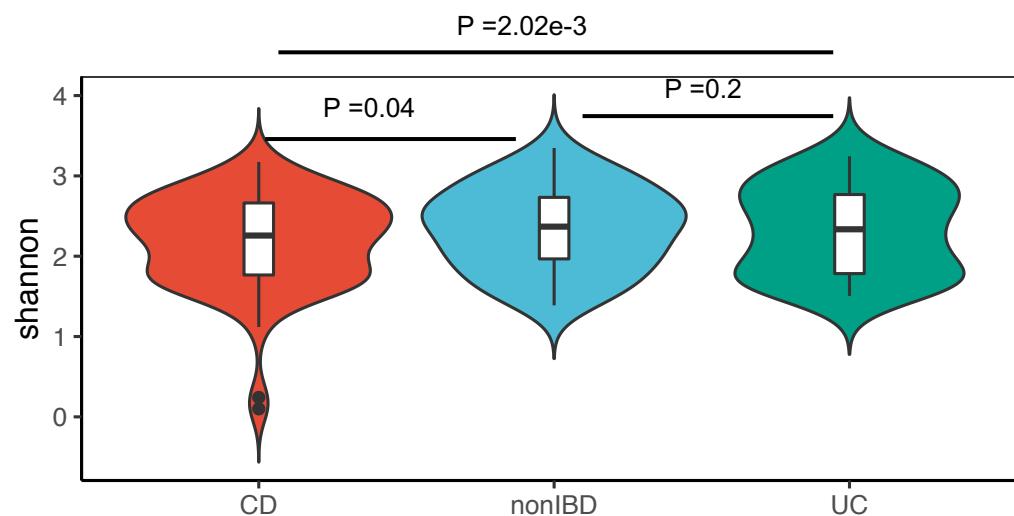
UC colon

UC ileum

CD/UC/nonIBD microbial patterns (HMP2 validation)

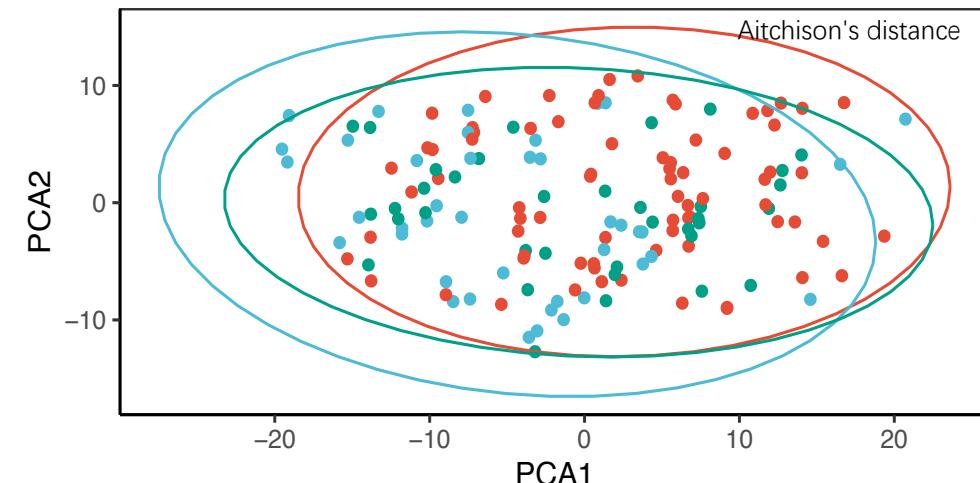
A

UC is more similar to nonIBD than CD

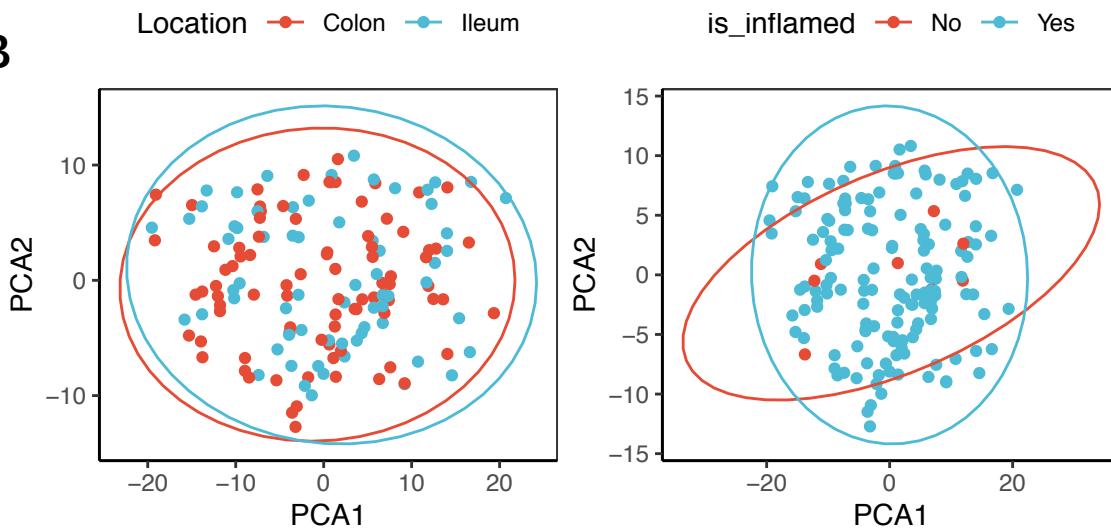


C

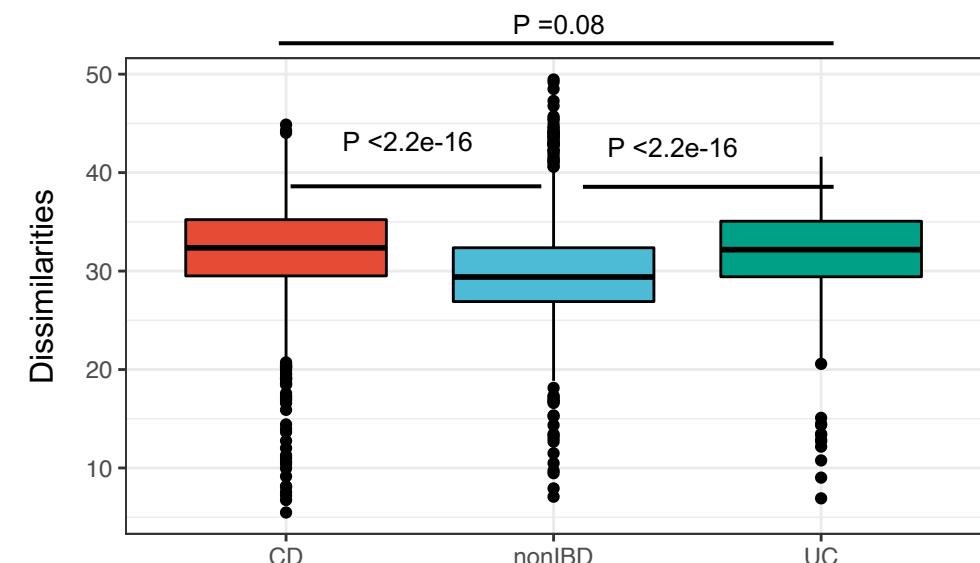
IBD has larger microbial variation than nonIBD

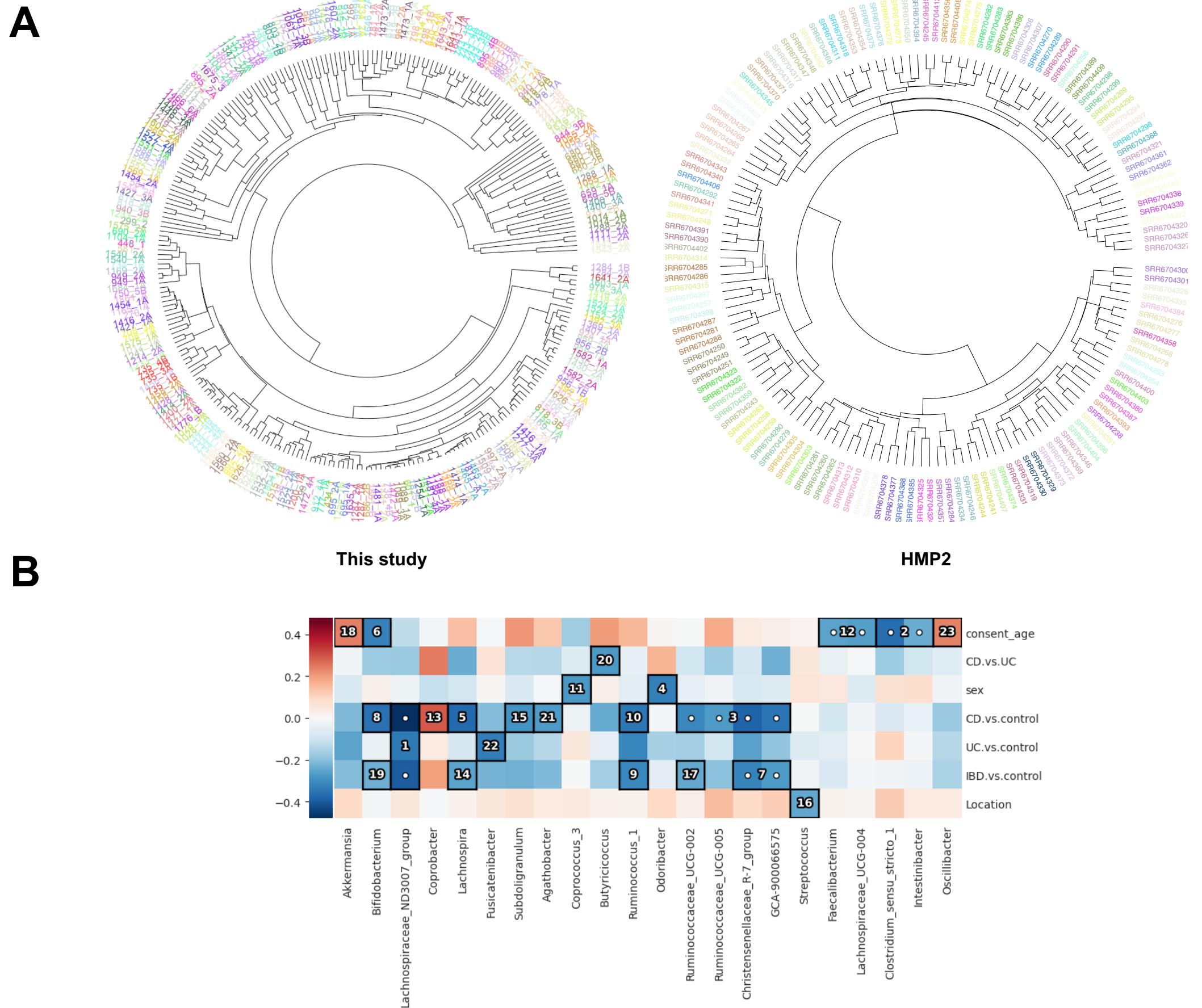


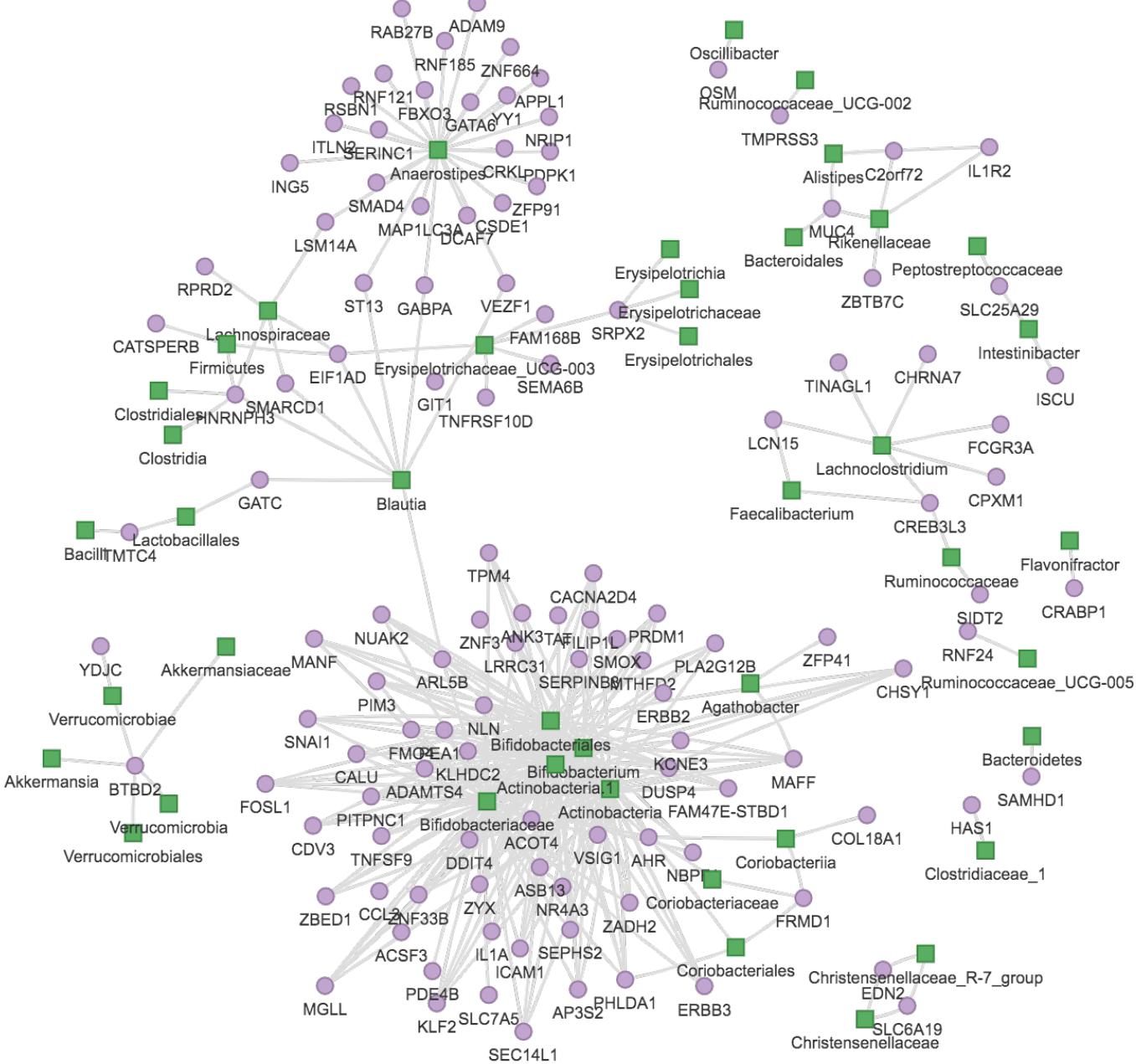
B

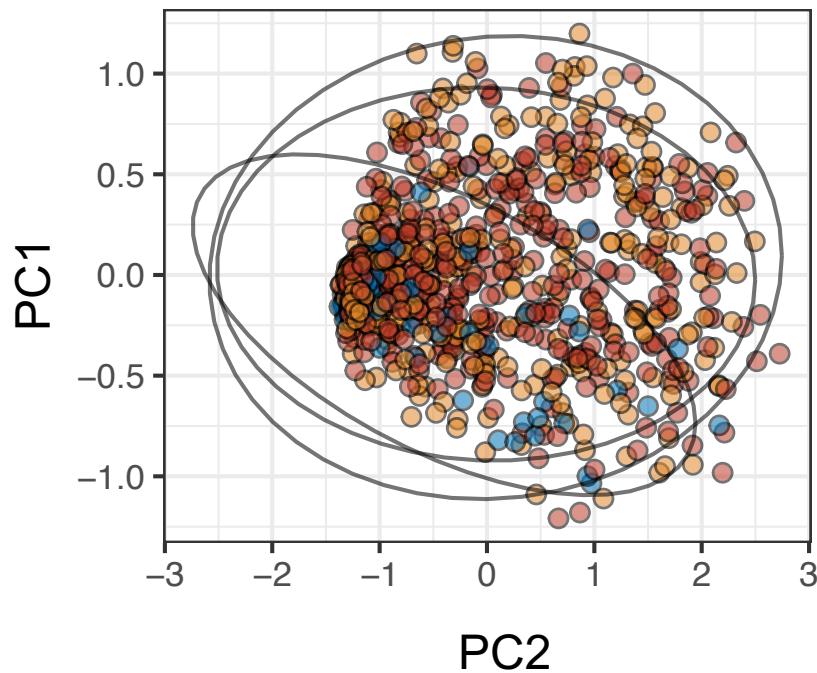


Dissimilarities

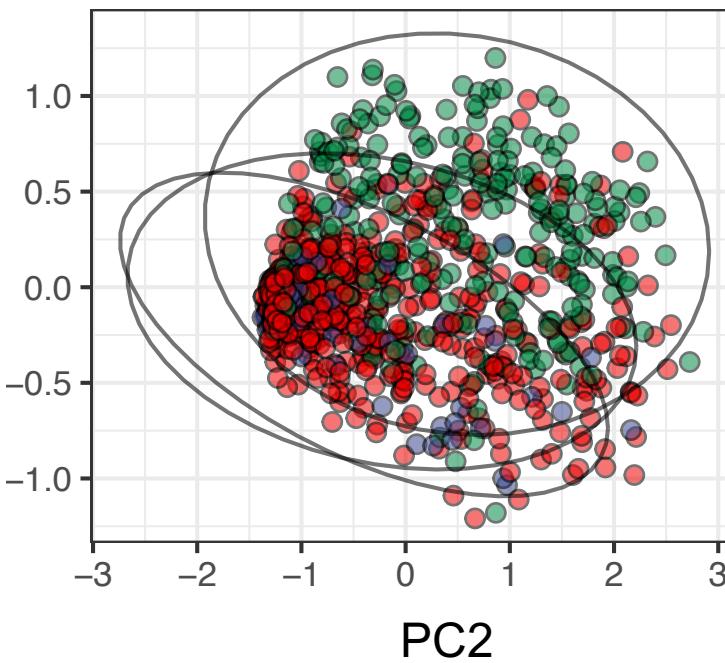




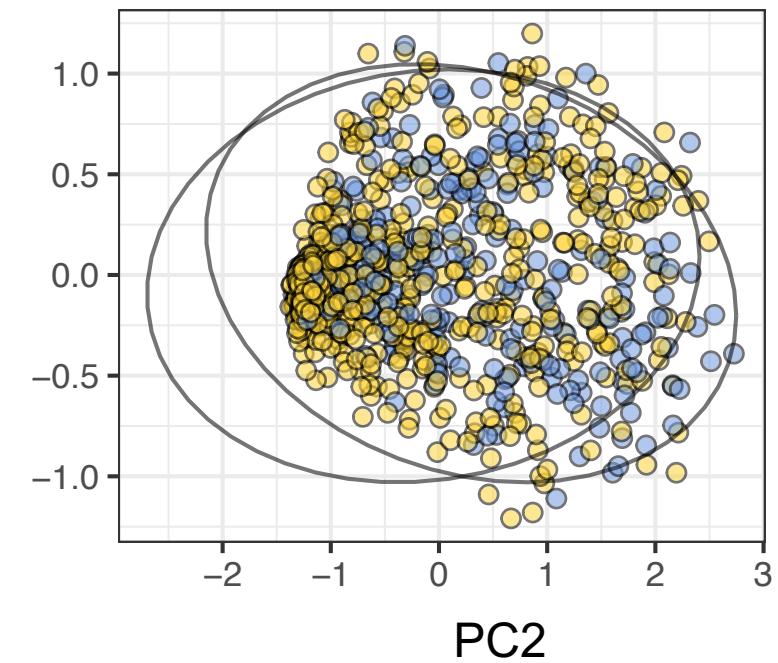




- UC
- Non-IBD
- CD



- IBD-Inflamed
- Non-IBD
- IBD-non-inflamed



- Ileum
- Colon