iScience, Volume *27*

Supplemental information

Gut-microbiota-based ensemble model predicts prognosis of pediatric inflammatory bowel disease

Sung Min Ha, Kihyun Lee, Gun-Ha Kim, Jakub Hurych, Ondřej Cinek, and Jung Ok Shim

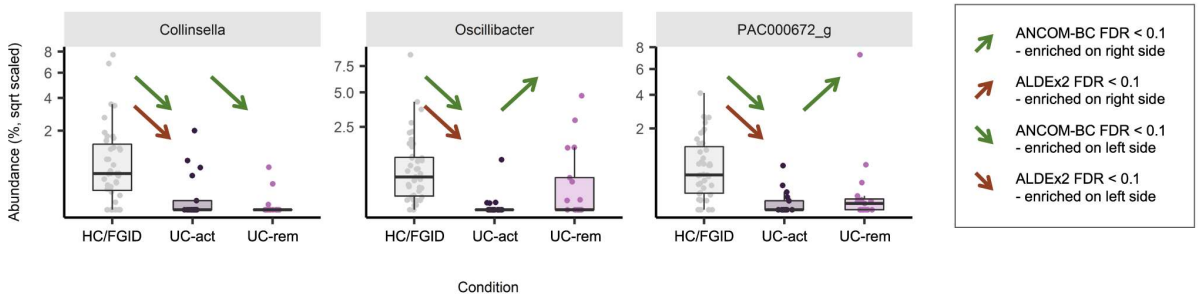

Figure S1: Bacterial taxa displaying differential abundance in association with UC, related to Figure 4. The differential abundance of taxa was tested at the genus level using ANCOM-BC and ALDEx2 using two different schemes: between the controls (including FGID) and UC-act; and between UC-act and UC-rem. The genera that showed significance (FDR < 0.1) in both schemes were considered associated with inflammation in UC and are shown in the figure. The boxplots represent interquartile range, along with the individual data points shown as dots. FGID, functional gastrointestinal diseases; UC, ulcerative colitis; UC-act, active UC; UC-rem, remission UC

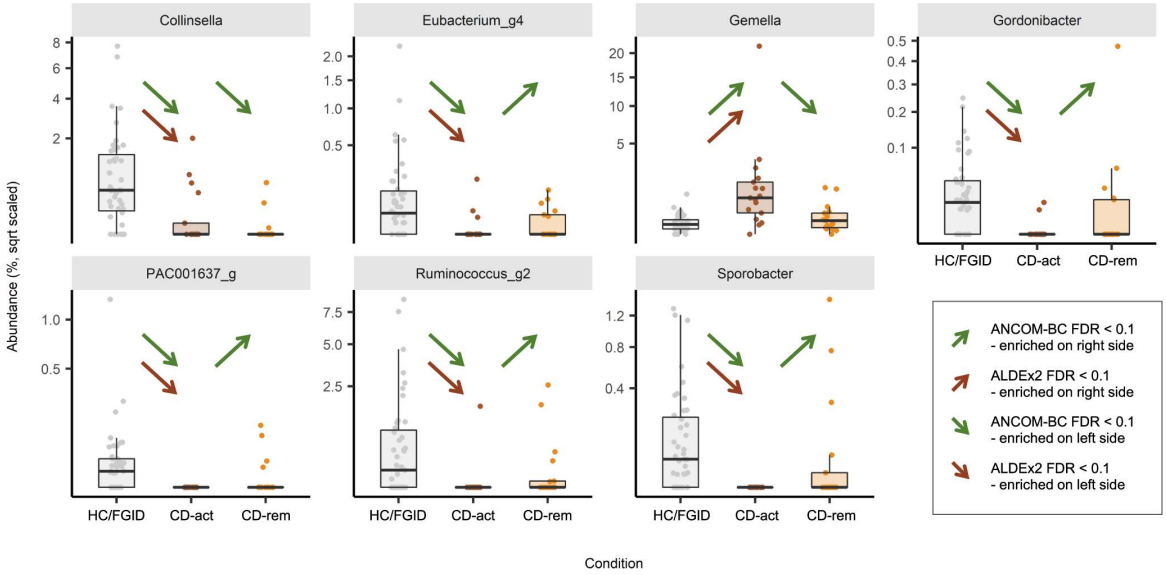

Figure S2: Bacterial taxa displaying differential abundance in association with CD, related to Figure 4. The differential abundance of the taxa was tested at the genus level using ANCOM-BC and ALDEx2 using two different schemes: between the controls (including FGID) and CD-act (active CD); and between CD-act and CD-rem (remission CD). The genera that showed significance (FDR < 0.1) in both schemes were considered associated with inflammation in CD and are shown in the figure. The boxplots represent interquartile range, along with the individual data points shown as dots. CD, Crohn's disease; CD-act, active CD; CD-rem, remission CD; FGID, functional gastrointestinal diseases; rem, remission

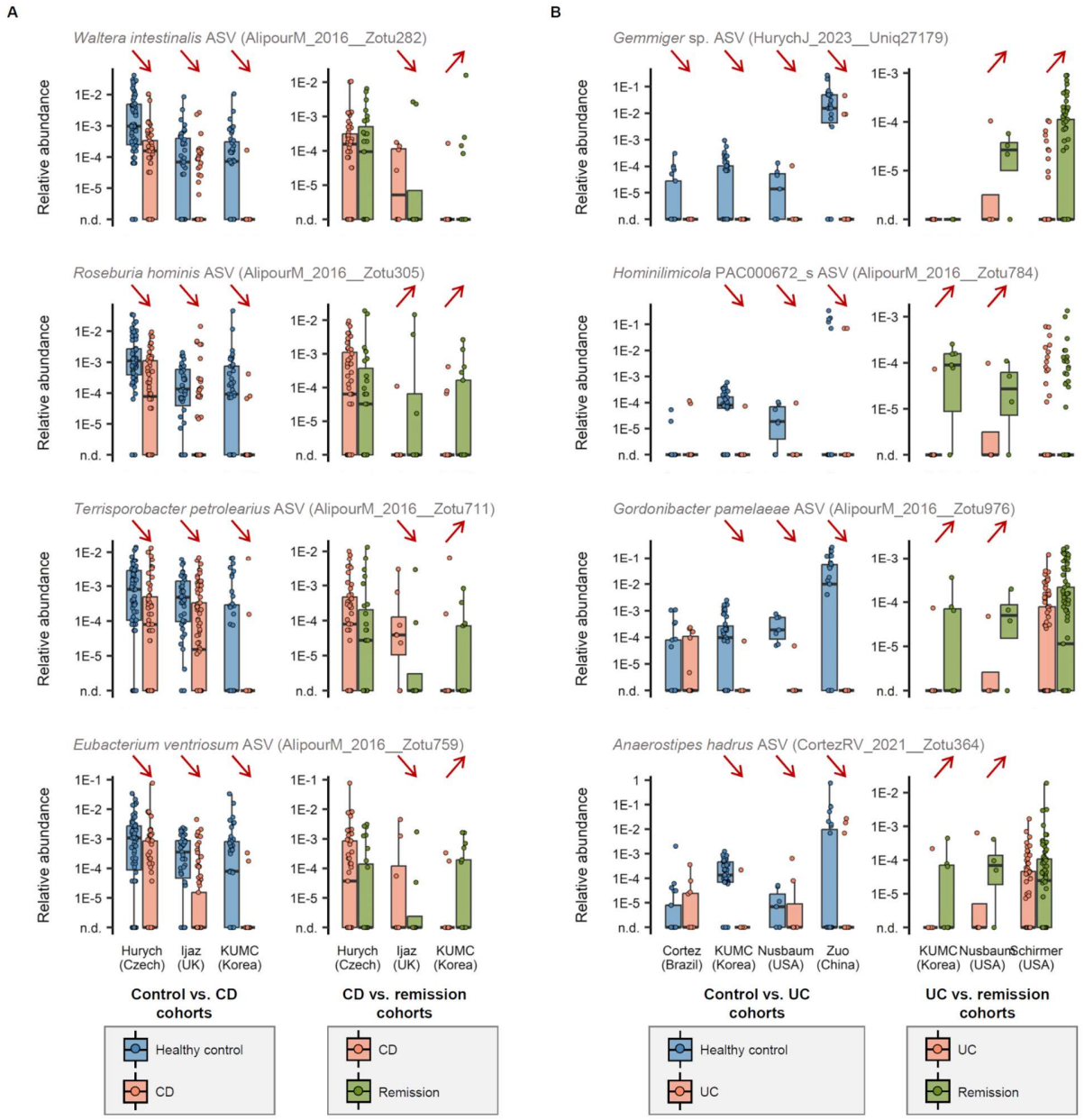

Figure S3: Taxonomic markers selected from differential abundance analysis of individual cohorts, related to Figure 4. A) The ASVs that were most frequently significant (FDR < 0.1) in the comparisons of healthy controls versus CD (n = 3) and CD versus remission (n = 3). B) The ASVs that were most frequently significant (FDR < 0.1) in the comparisons of heathy controls versus UC (n = 4) and UC versus remission (n =3). All plots show the ASV's relative abundance in log scale on the y axis, and the boxplots represent interquartile range, along with the individual data points shown as dots. Slanted arrows above the box plots of each cohort were placed only when FDR < 0.1, and the direction of the arrows indicate whether the ASV was depleted or enriched in the group displayed on the right side.

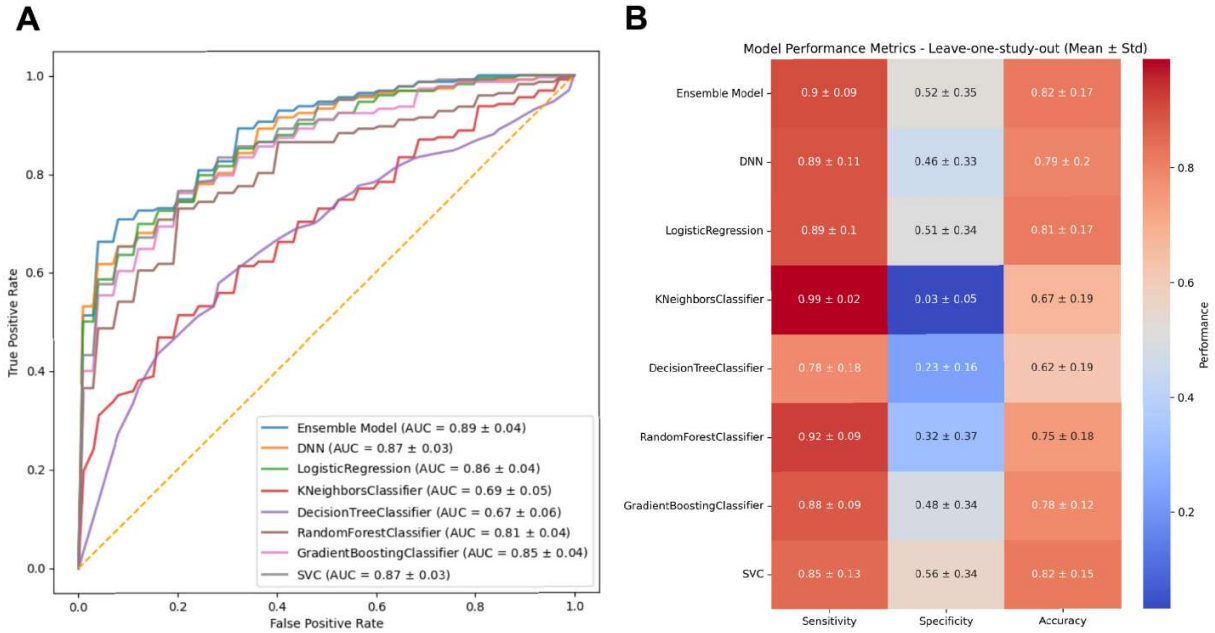

Figure S4: Comparison of performance various models trained with patients' future remission states based on the microbiota profile sampled during the active disease state measured with leave-one-study-out method, related to Figure 5. A) Receiver operating characteristic curves of the eight machine learning models trained on the baseline microbiota 16S ASV read counts and clinical features to predict patient remission. B) Comparison of classification performance of the eight machine learning models tested based on sensitivity, specificity, and accuracy metrics. ASVs, amplicon sequence variants.

Table S1: Alpha diversity comparison statistics calculated from the Korean cohort, related to Figure 2.

Comparison	Fold-difference in median Chao1	P-value (Wilcoxon test)	FDR
Control vs. FGID	1.01	0.64	0.64
Control vs. CD-act	1.43	8.5E-6	5.1^{E-5}
Control vs. CD-rem	1.15	0.013	0.026
FGID vs. CD-act	1.47	1.0E-4	3.0^{E-4}
FGID vs. CD-rem	1.19	0.035	0.053
CD-act vs. CD-rem	1.24	0.092	0.11
Control vs. UC-act	1.31	0.043	0.19
Control vs. UC-rem	1.05	0.91	0.99
FGID vs. UC-act	1.36	0.073	0.19
FGID vs. UC-rem	1.01	0.78	0.99
UC-act vs. UC-rem	1.37	0.097	0.19
New-onset vs. recurrent (exacerbation)	1.26	0.20	0.24
New-onset vs. CR	1.18	0.075	0.15
New-onset vs. MR	1.15	0.19	0.24
Recurrent (exacerbation) vs. CR	1.48	0.014	0.087
Recurrent (exacerbation) vs. MR	1.45	0.044	0.13
CR vs. MR	1.02	0.91	0.13

CD-act, Crohn's disease active inflammatory state; CD-rem, Crohn's disease remission state; CR, clinical remission; FDR, false discovery rate; FGID, functional gastrointestinal disorder; MR, mucosal remission

Table S7: Literature search results of known host–microbe interaction attributes of the taxa represented by the top 20 ASVs with highest SHAP scores in our Ensemble model, related to Figure 6.

Taxonomy (at the lowest resolved rank)	Known attribute of the taxon in host–microbe interaction	Reference
Actinobacteria *Bifidobacterium adolescentis*	Known as one of the primary producers of acetate among gut bacteria. Forms a cross-feeding relationship with butyrate-producing gut bacteria. Highly prevalent in adults rather than children.	O'Riordan et al., Rivière et al. [S1,2]
Bacteroidetes *Bacteroides* sp.	This ASV was not close enough to any of the known *Bacteroides* spp. hence remained unclassified at species level. *Bacteroides* in general, are one of the most abundant bacterial genera in the normal gut microbiome. They are primary producers of acetate and propionate and capable of utilizing various dietary polysaccharides.	Zafar and Saier [S3]
Firmicutes *Streptococcus*	*Streptococcus* spp. reportedly have a significant association with CD and UC. Some studies proposed that an increased abundance of common oral bacteria such as streptococci in the gut microbiota is associated with systemic inflammatory markers.	Ocansey et al., Sayols-Baixeras et al. [S4,5]
Firmicutes *Anaerobutyricum soehngenii*	The species was originally described as a butyrate- and propionate-producing bacterium discovered in infant gut. Supplementation of this species led to amelioration of metabolic syndrome in mouse studies as well as in randomized clinical trials on human subjects.	Shetty et al., Koopen et al., Attaye et al. [S6,7,8]
Firmicutes *Enterocloster* sp.	This ASV is yet undescribed species of *Enterocloster*. Known species in *Enterocloster* have been reported to be associated with autism-spectrum disorder (*Enterocloster bolteae*), and also to elicit gut mucosal immune response resulting in protective effect against *Salmonella* infection (*Enterocloster clostridioformis*).	Pequegnat et al., Beresford-Jones et al.[S9,10]
Firmicutes *Hominiventricola* sp.	This ASV is a yet undescribed species of *Hominiventricola*. The genus *Hominiventricola* is little studied, with only a single recently described species and no appearance in literatures.	
Firmicutes *Ruminococcus lactaris*	Little evidence has been produced regarding physiological functions of *Ruminococcus lactaris* through *in vitro* studies, but the species has been associated with diseases such as autism-spectrum disorder. A closely related and extremely actively studied species, *Ruminococcus gnavus*, is deeply involved in gut microbiome mediation of human health and diseases.	Dan et al., Crost et al.[S11,12]
Firmicutes *Faecalibacterium duncaniae*	*Faecalibacterium duncaniae* exhibited anti-inflammatory effects in numerous animal model studies. Strains of this species produce butyrate, along with several other putative mechanisms of host immune regulation.	Martín et al. [S13]
Firmicutes *Peptostreptococcus anaerobius*	*Peptostreptococcus anaerobius* is known as an oral pathogen, whose enrichment in gut microbiome was shown to exacerbate colorectal cancer and impairs the response to anti-PD1 therapy, with known mechanistic evidence for the bacterium's immunomodulatory functions. The same species is also known as a major producer of propionate in infant gut microbiome.	Liu et al., Laursen et al. [S14,15]
Firmicutes *Veillonella* sp.	This ASV is yet undescribed species of *Veillonella.* The genus includes *Veillonella parvula,* anaerobic nonsaccharolytic oral microorganisms that rely on organic acid metabolism for energy and exhibit heightened proliferation within the intestinal milieu of individuals diagnosed with IBD, and *Veillonella dispar,* which has been described multiple times as a strong marker in IBD.	Rojas-Tapias et al., Altomare et al., Schirmer et al. [S16,17,19]
Firmicutes *Megasphaera micronuciformis* and *Megasphaera* sp.	*Megasphaera micronuciformis* was initially described with isolates from patients with infection, i.e., anaerobic culture on blood agar. Little is known about this species. The genus *Megasphaera* has been described to be statistically associated with numerous disease conditions including pediatric IBD.	Marchandin et al., El Mouzan et al. [S20,21]
Fusobacteria *Fusobacterium*	A *Fusobacterium pseudoperiodonticum* ASV and a species-unresolved *Fusobacterium* ASV were included in the top 20 ASVs. *Fusobacterium* spp. have been pointed as significant markers of IBD.	Meade et al. [S22]
Proteobacteria *Haemophilus*	*Haemophilus* species were previously shown to increase in CD. It is considered a common oral taxa.	Gevers et al., Elmaghrawy et al. [S23,24]

ASVs, amplicon sequence variants; CD, Crohn's disease; IBD, inflammatory bowel disease; SCFAs, short-chain fatty acids; SHAP, Shapley Additive exPlanation; sIgA, secretory immunoglobulin A; UC, ulcerative colitis

Table S9: Results from hyperparameter tuning during the prognostic model training, related to STAR Methods.

Models	Parameters tested	Range	Best parameter
DNN	l1_value	[0.001, 0.01, 0.1]	0.001
DNN	l2_value	[0.001, 0.01, 0.1]	0.001
DNN	dropout_rate	[0.3, 0.5, 0.7]	0.3
DNN	hidden_layers	[1,2,3,4,5,6,7,8,9,10]	2
DNN	units	range(10,200)	99
DNN	epochs	[50, 100, 150, 200]	50
DNN	batch_size	[16, 32, 64, 128]	128
Logistic Regression	C (Inverse of regularization strength)	[0.01, 0.1, 1, 10, 100]	0.1
K-nearest neighbor	n_neighbors	range(1, 30)	8
K-nearest neighbor	weights	[uniform, distance]	distance
K-nearest neighbor	p (Power parameter for the Minkowski metric)	[1, 2]	1
Decision Tree	max_depth	range(1, 20)	15
Decision Tree	min_samples_split	range(2, 20)	2
Decision Tree	min_samples_leaf	range(1, 20)	10
RandomForest	n_estimators	[100, 200, 300, 400, 500]	400
RandomForest	max_features	[sqrt, log2, None]	sqrt
RandomForest	max_depth	range(1, 20)	18
RandomForest	min_samples_split	range(2, 20)	4
RandomForest	min_samples_leaf	range(1, 20)	2
GradientBoosting	n_estimators	[100, 200, 300, 400, 500]	300
GradientBoosting	learning_rate	[0.01, 0.05, 0.1, 0.2]	0.2
GradientBoosting	subsample	[0.5, 0.7, 1.0]	0.5
GradientBoosting	max_depth	range(1, 20)	19
SVC	C (Regularization parameter)	[0.01, 0.1, 1, 10, 100]	0.1
SVC	kernel	[linear, poly, rbf, sigmoid]	linear
SVC	gamma	[scale, auto]	scale

Supplemental References

[S1]. O'Riordan K.J., Collins M.K., Moloney G.M., Knox E.G., Aburto M.R., Fülling C., Morley S.J., Clarke G., Schellekens H., and Cryan J.F. (2022). Short chain fatty acids: Microbial metabolites for gut-brain axis signalling. Mol. Cell. Endocrinol. 546, 111572. 10.1016/j.mce.2022.111572.

[S2]. Rivière, A., Selak, M., Lantin, D., Leroy, F., and De Vuyst, L. (2016). Bifidobacteria and butyrate-producing colon bacteria: Importance and strategies for their stimulation in the human gut. Front. Microbiol. 7, 979. 10.3389/fmicb.2016.00979.

[S3]. Zafar, H. and Saier, M.H. Jr. (2021). Gut *Bacteroides* species in health and disease. Gut Microbes 13(1), 1-20. 10.1080/19490976.2020.1848158.

[S4]. Wiredu Ocansey, D.K., Hang, S., Yuan, X., Qian, H., Zhou, M., Valerie Olovo, C., Zhang, X., and Mao, F. (2023). The diagnostic and prognostic potential of gut bacteria in inflammatory bowel disease. Gut Microbes 15, 2176118. 10.1080/19490976.2023.2176118.

[S5]. Sayols-Baixeras, S., Dekkers, K.F., Baldanzi, G., Jönsson, D., Hammar, U., Lin, Y.T., Ahmad, S., Nguyen, D., Varotsis, G., Pita, S., et al. (2023). *Streptococcus* species abundance in the gut is linked to subclinical coronary atherosclerosis in 8973 participants from the SCAPIS cohort. Circulation 148, 459–472. 10.1161/CIRCULATIONAHA.123.063914.

[S6]. Shetty, S.A., Zuffa, S., Bui, T.P.N., Aalvink, S., Smidt, H., De Vos, W.M. (2018) Reclassification of *Eubacterium hallii* as *Anaerobutyricum hallii* gen. nov., comb. nov., and description of *Anaerobutyricum soehngenii* sp. nov., a butyrate and propionate-producing bacterium from infant faeces. Int. J. Syst. Evol. Microbiol. 68(12), 3741-3746. 10.1099/ijsem.0.003041.

[S7]. Koopen, A., Witjes, J., Wortelboer, K., Majait, S., Prodan, A., Levin, E., Herrema, H., Winkelmeijer, M., Aalvink, S., Bergman, J.J.G.H.M., et al. (2022). Duodenal *Anaerobutyricum soehngenii* infusion stimulates GLP-1 production, ameliorates glycaemic control and beneficially shapes the duodenal transcriptome in metabolic syndrome subjects: a randomised double-blind placebo-controlled cross-over study. Gut 71(8), 1577-1587. 10.1136/gutjnl-2020-323297.

[S8]. Attaye, I., Witjes, J.J., Koopen, A.M., van der Vossen, E.W.J., Zwirs, D., Wortelboer, K., Collard, D., Kemper, E.M., Winkelmeijer, M., Holst, J.J., et al. (2024). Oral *Anaerobutyricum soehngenii* augments glycemic control in type 2 diabetes. iScience 27(8), 110455. 10.1016/j.isci.2024.110455.

[S9]. Pequegnat, B., Sagermann, M., Valliani, M., Toh, M., Chow, H., Allen-Vercoe, E., and Monteiro, M.A. (2013). A vaccine and diagnostic target for *Clostridium bolteae*, an autism-associated bacterium. Vaccine. 31(26), 2787-90. 10.1016/j.vaccine.2013.04.018.

[S10]. Beresford-Jones, B.S., Suyama. S., Clare. S., Soderholm, A., Xia, W., Sardar, P., Harcourt, K., Lawley, T.D., and Pedicord, V.A. (2023). *Enterocloster clostridioformis* induces host intestinal epithelial responses that protect against *Salmonella* infection. bioRxiv 2023.07.20.549886. 10.1101/2023.07.20.549886.

[S11]. Dan, Z., Mao, X., Liu, Q., Guo, M., Zhuang, Y., Liu, Z., Chen, K., Chen, J., Xu, R., Tang, J., et al. (2020). Altered gut microbial profile is associated with abnormal metabolism activity of Autism Spectrum Disorder. Gut Microbes 11(5), 1246-1267. 10.1080/19490976.2020.1747329.

[S12]. Crost, E.H., Coletto, E., Bell, A., and Juge, N. (2023). *Ruminococcus gnavus*: friend or foe for human health. FEMS. Microbiol. Rev. 47(2), fuad014. 10.1093/femsre/fuad014.

[S13]. Martín, R., Rios-Covian, D., Huillet, E., Auger, S., Khazaal, S., Bermúdez-Humarán, L.G., Sokol, H., Chatel, J.M., and Langella, P. (2023). *Faecalibacterium*: a bacterial genus with promising human health applications. FEMS Microbiol. Rev. 47(4), fuad039. 10.1093/femsre/fuad039.

[S14]. Liu, Y., Wong, C.C., Ding, Y., Gao, M., Wen, J., Lau, H.C., Cheung, A.H., Huang, D., Huang, H., and Yu, J. (2024). *Peptostreptococcus anaerobius* mediates anti-PD1 therapy resistance and exacerbates colorectal cancer via myeloid-derived suppressor cells in mice. Nat. Microbiol. 9(6), 1467-1482. 10.1038/s41564-024-01695-w.

[S15]. Laursen, M.F., Sinha, A.K., Pedersen, M., and Roager, H.M. (2023). Key bacterial taxa determine longitudinal dynamics of aromatic amino acid catabolism in infants' gut. Gut Microbes 15(1), 2221426. 10.1080/19490976.2023.2221426.

[S16]. Rojas-Tapias, D.F., Brown, E.M., Temple, E.R., Onyekaba, M.A., Mohamed, A.M.T., Duncan, K., Schirmer, M., Walker, R.L., Mayassi, T., Pierce, K.A., et al. (2022). Inflammation-associated nitrate facilitates ectopic colonization of oral bacterium *Veillonella parvula* in the intestine. Nat. Microbiol. 7, 1673–1685. 10.1038/s41564-022-01224-7.

[S17]. Altomare, A., Putignani, L., Del Chierico, F., Cocca, S., Angeletti, S., Ciccozzi, M., Tripiciano, C., Dalla Piccola, B., Cicala, M., and Guarino, M.P.L. (2019). Gut mucosal-associated microbiota better discloses inflammatory bowel disease differential patterns than faecal microbiota. Dig. Liver Dis. 51, 648–656. 10.1016/j.dld.2018.11.021.

[S18]. Schirmer, M., Denson, L., Vlamakis, H., Franzosa, E.A., Thomas, S., Gotman, N.M., Rufo, P., Baker, S.S., Sauer, C., Markowitz, J., et al. (2018). Compositional and temporal changes in the gut microbiome of pediatric ulcerative colitis patients are linked to disease course. Cell Host Microbe 24, 600–610.e4. 10.1016/j.chom.2018.09.009.

[S20]. Marchandin, H., Jumas-Bilak, E., Gay, B., Teyssier, C., Jean-Pierre, H., Siméon de Buochberg, M., Carrière, C., and Carlier, J.P. (2003). Phylogenetic analysis of some Sporomusa sub-branch members isolated from human clinical specimens: description of *Megasphaera micronuciformis* sp. nov. Int. J. Syst. Evol. Microbiol. 53, 547–553. 10.1099/ijs.0.02378-0.

[S21]. El Mouzan, M.I., Winter, H.S., Assiri, A.A., Korolev, K.S., Al Sarkhy, A.A., Dowd, S.E., Al Mofarreh, M.A., and Menon, R. (2018). Microbiota profile in new-onset pediatric Crohn's disease: data from a non-Western population. Gut Pathog. 10, 49. 10.1186/s13099-018-0276-3.

[S22]. Meade, S., Liu Chen Kiow, J., Massaro, C., Kaur, G., Squirell, E., Bressler, B., and Lunken, G. (2023). Gut microbiome-associated predictors as biomarkers of response to advanced therapies in inflammatory bowel disease: a systematic review. Gut Microbes 15(2), 2287073. 10.1080/19490976.2023.2287073.

[S23]. Gevers, D., Kugathasan, S., Denson, L.A., Vázquez-Baeza, Y., Van Treuren, W., Ren, B., Schwager, E., Knights, D., Song, S.J., Yassour, M., et al. (2014). The treatment-naive microbiome in new-onset Crohn's disease. Cell Host Microbe 15, 382–392. 10.1016/j.chom.2014.02.005.

[S24]. Elmaghrawy, K., Fleming, P., Fitzgerald, K., Cooper, S., Dominik, A., Hussey, S., et al. (2023). The Oral Microbiome in Treatment-Naïve Paediatric IBD Patients Exhibits Dysbiosis Related to Disease Severity That Resolves Following Therapy. J Crohn Colitis 17, 553–564. 10.1093/ecco-jcc/jjac155.