Supplemental Materials

The Presence of Periodontal Pathogens in Gastric Cancer

Marcel A. de Leeuw & Manuel X. Duval, GeneCreek

Contents

List of Figures

List of Tables

Revised manuscript

Supplemental materials

For the purpose of comparison, we included in our analysis five colorectal cancer (CRC) mucosa biopsy data sets we had previously analyzed, Table S1.

Table S1: Colorectal cancer biopsy samples used in this study. n: number of samples used, 16S: variable regions covered.

BioProject	SRA	n	16S	region
PRJEB6070	ERP005534	96	V4	Germany
PRJNA298957	SRP064975	98	V3-V4	China, Shanghai
PRJNA325650	SRP076561	50	V3-V4	Malaysia
PRJNA404030	SRP117763	29	V3-V4	New Zealand
PRJNA445346	SRP137015	211	V3-V5	U.S.A.
total		484		

The following large scale 16S microbiome studies, in addition to the studies listed in main manuscript Table 1, were used in the co-exclusion analysis of the gastric cancer species, Table S2.

Table S2: Studies used for co-exclusion analysis. 16S: variable regions sequenced.

SRA	16S	samples	publication
SRP062005	V4	stool	[1]
SRP076743	V4	stool	[2]
ERP009494	V3-V4	stool	[3]
SRP040765	V4	lower gut biopsies and stool	[4]
SRP104731	V4	stool	[5]
ERP006339	V4	stool	[6]
SRP070848	V4	sputum	[7]
SRP060025	V4-V5	sputum	[8]
SRP043334	V1-V2	sputum	[9]
SRP087648	V4	stool	[10]
SRP064975	V3-V4	lower gut biopsies	[11]
ERP012803	V4	stool	[12]
SRP063707	V1-V3	skin	[13]
SRP097785	V4	stool	[14]
SRP068187	V4	serum	[15]
SRP068473	V4	stool	[16]
SRP076281	V4	stool	[17]
SRP077299	V4	sputum	[18]
ERP013984	V4	eye and skin	[19]
SRP057700	V1-V2	stool	[20]
SRP090628	V4	stool	[21]

microbial community types

Using Dirichlet Multinomial Mixtures on the combined relative abundances of the nine datasets (n=1,544) listed in main manuscript Table 1, we obtain an optimal goodness of fit at k=5 communities according to the Laplace and AIC evaluations, figure S1. The breakdown of samples from the various datasets along the community types

is given in Table S3.

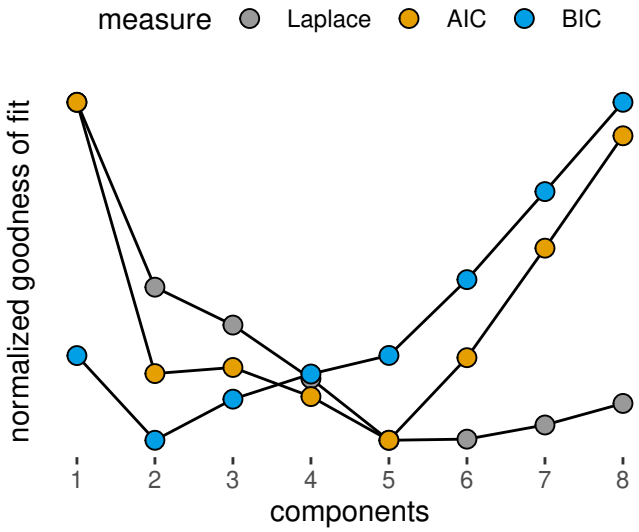

Figure S1: Goodness of fit of the DMM models at different k.

Table S3: Distribution of community types across studies. The five community types are in columns.

study	dmm 1	dmm 2	dmm 3	dmm 4	dmm 5
ERP023334	10	30		81	
ERP023753	16	5		13	
ERP024440	1	13		18	
SRP070925	2				117
SRP128749	635	34			
SRP154244		83	179	39	
SRP165213	23	9			
SRP172818	155	17		1	
SRP200169	42	21			

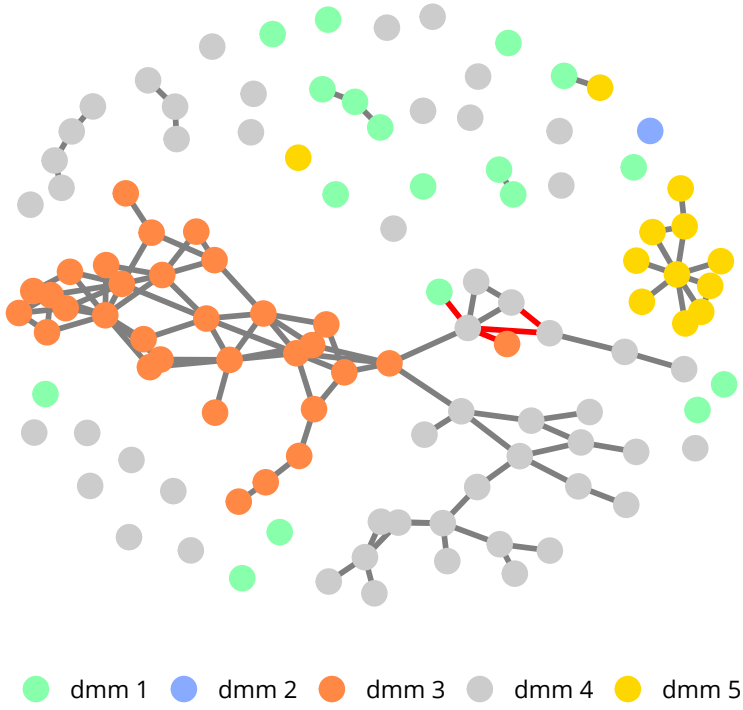

Figure S2: Interaction network between species relevant for community types. The top 100 species relevant for distinction between the five community types are displayed.

The first two community types are dominated by a few species mostly without interaction. The majority of healthy donor samples was located in community type one, together with certain tumor samples.

Of note, community types three and five received contributions from a single study each, Table S3. Hence, although we find multinomial mixtures and inverse covariance networks were in good agreement for overall gastric microbiota composition, we observed potentially only a subset of regionally or otherwise determined gastric microbiota. Among the top 100 differentiating species we found 62 distinct genera, further highlighting the diversity. Table S4 lists the 18 genera with more than one species.

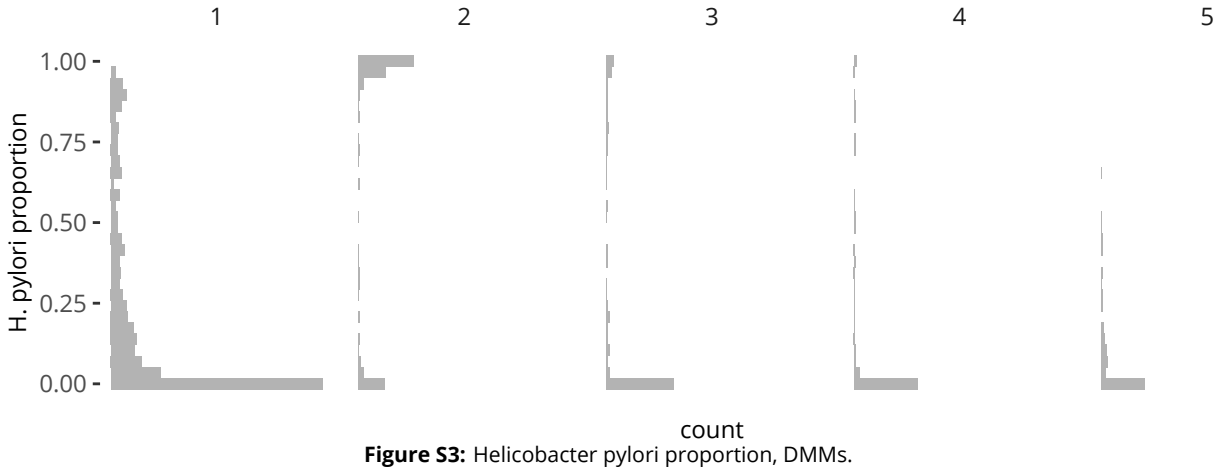

Figure S3: Helicobacter pylori proportion, DMMs.

Table S4: Gastric mucosa genera. Only genera with more than one species are listed.

genus	species
Prevotella	10
Streptococcus	9
Acinetobacter	4
Campylobacter	4
Porphyromonas	4
Arthrobacter	3
Fusobacterium	3
Leuconostoc	3
Methylobacterium	3
Sphingomonas	3
Veillonella	3
Actinomyces	2
Alloprevotella	2
Bacillus	2
Brevundimonas	2
Clostridium	2
Haemophilus	2
Lactococcus	2
Neisseria	2

Further indication that the DMMs are distinct in nature can be found in the projection of alpha diversity, using the phylogenetic diversity (whole tree), figure S4. The *Helicobacter pylori* dominated community type three has

the lowest diversity.

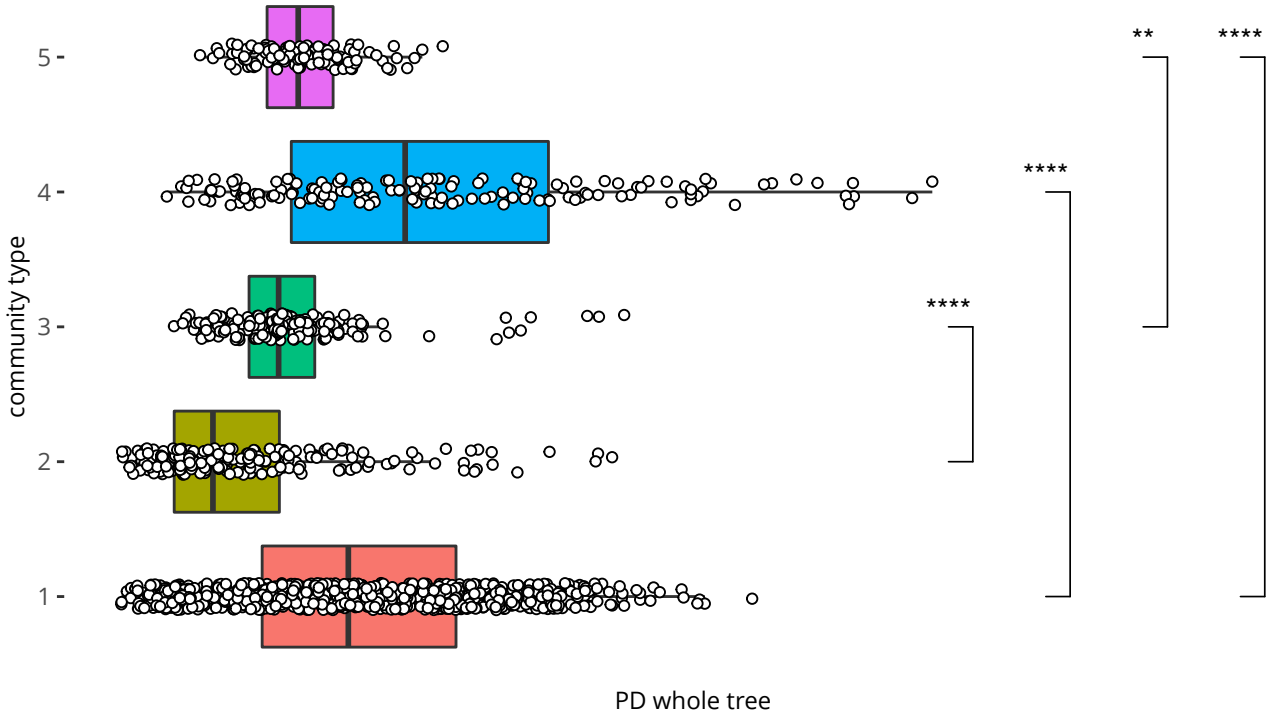

Figure S4: Alpha diversity of community types. Phylogenetic diversity (whole tree).

anatomic location

Data set SRP154244 presents samples from different gastric locations in patients with gastritis, intestinal metaplasia and gastric cancer. We investigate if microbial signatures differ per anatomic location by training an RF model on two thirds of the samples and evaluating the model on the remaining third. Table S5 suggests the antral is well differentiated from the antrum and body, but the latter two are not differentiated. Thus at first sight, gastric location could at least in part explain differences in community types.

Table S5: RF classification of sampling location. Predictions are in columns. Multiclass AUC:0.788

location	antral	antrum	body
antral	72	3	0
antrum	6	11	0
body	1	6	1

To shine further light on this matter, we group corpus and antrum samples together and retrain an RF model on the whole of the SRP154244 dataset, retrieve differentiating species and build a SPIEC-EASI network, figure S5. Although we find significant separation between the two locations, especially when considering the negative correlations (in red), the separation is not as strict as the separation between community types. So it does not seem we can explain the distribution of datasets over the community typrs by difference in anatomic location alone. Of note, we find three bacteria encountered in colorectal cancer, *Fusobacterium nucleatum*, *Parvimonas micra*, *Peptostreptococcus stomatis* in interaction and associated with the corpus/antrum. *Helicobacter pylori* is

more abundant in the antral and not in interaction with any other species.

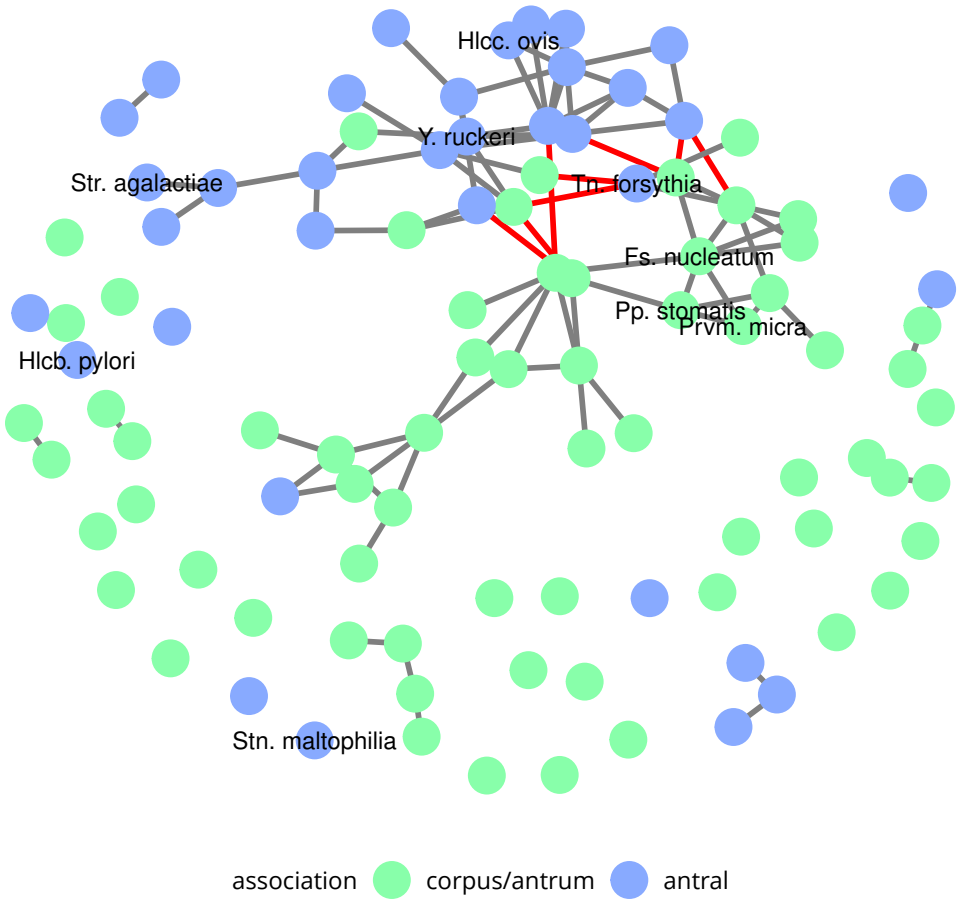

Figure S5: Interaction network between species relevant for gastric location. The top 100 species relevant for distinction between the two gastric locations are displayed. Opportunistic pathogens are labelled.

disease progress

Figure S6 provides a multidimensional scaling plot of the combined disease progress data sets SRP200169 (healthy subjects) and SRP070925. The corresponding Shannon species diversity distribution is geven in Figure S7. Shannon species diversity was also computed for disease progress data set ERP023334, Figure S8. Gastritis is characterized by dysbiosis as compared to healthy tissue, with a trend to reach normal diversity along the disease progress.

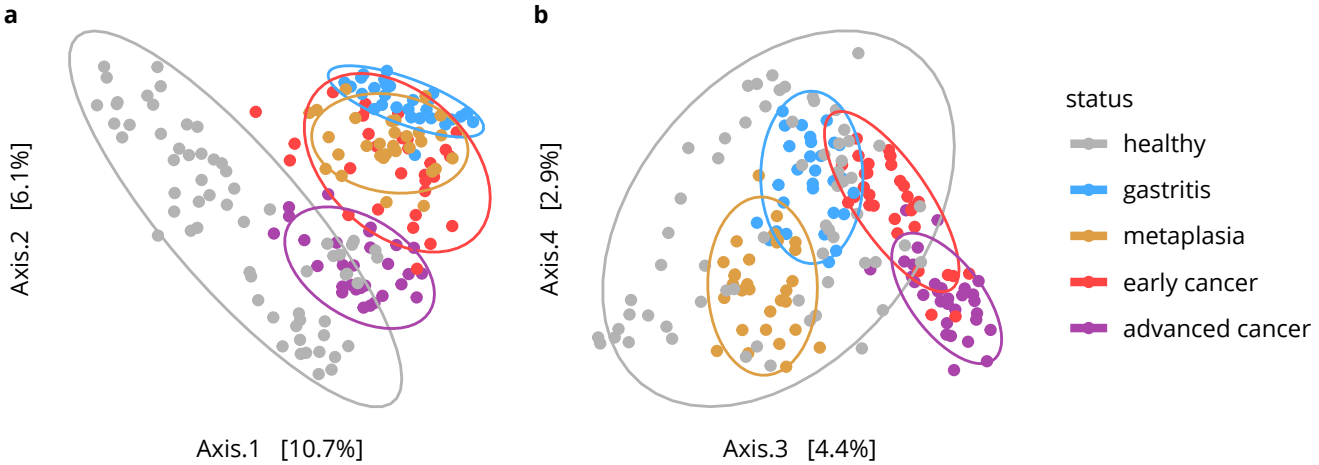

Figure S6: Multi-dimensional scaling of the disease progress data set. Unweighted UniFrac of ASVs is used as the distance metric.

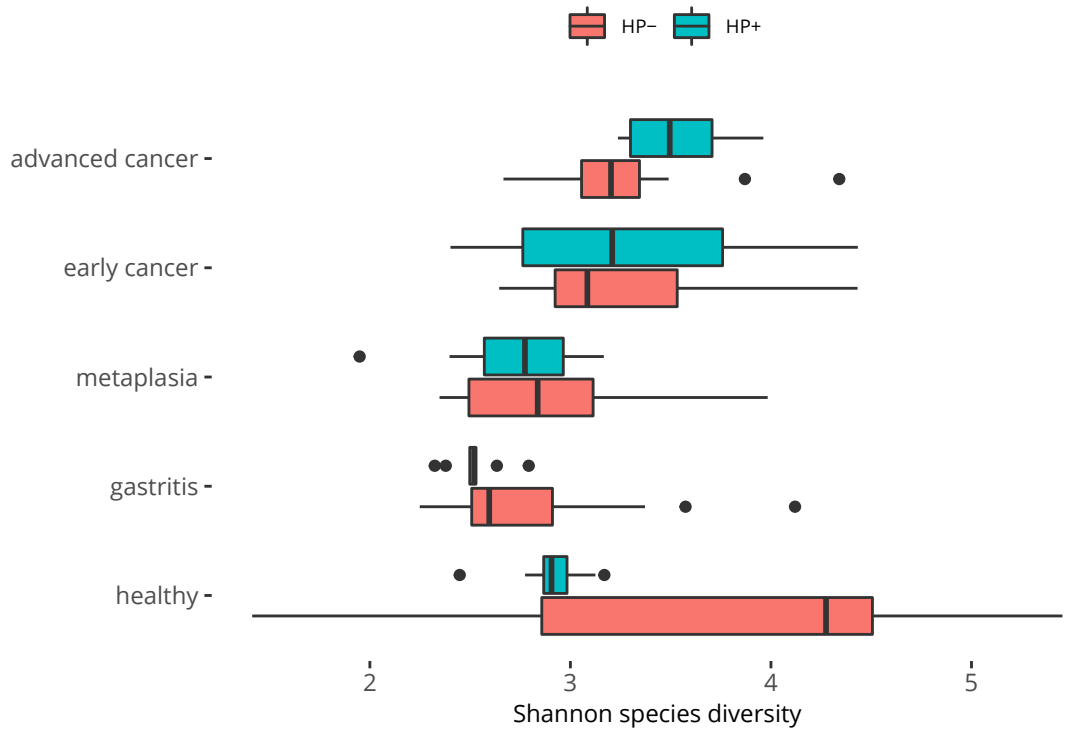

Figure S7: Shannon species diversity and disease progress, (SRP200169 + SRP070925). Helicobacter pylori positive (Hp+) and negative (Hp-) samples are distinguished.

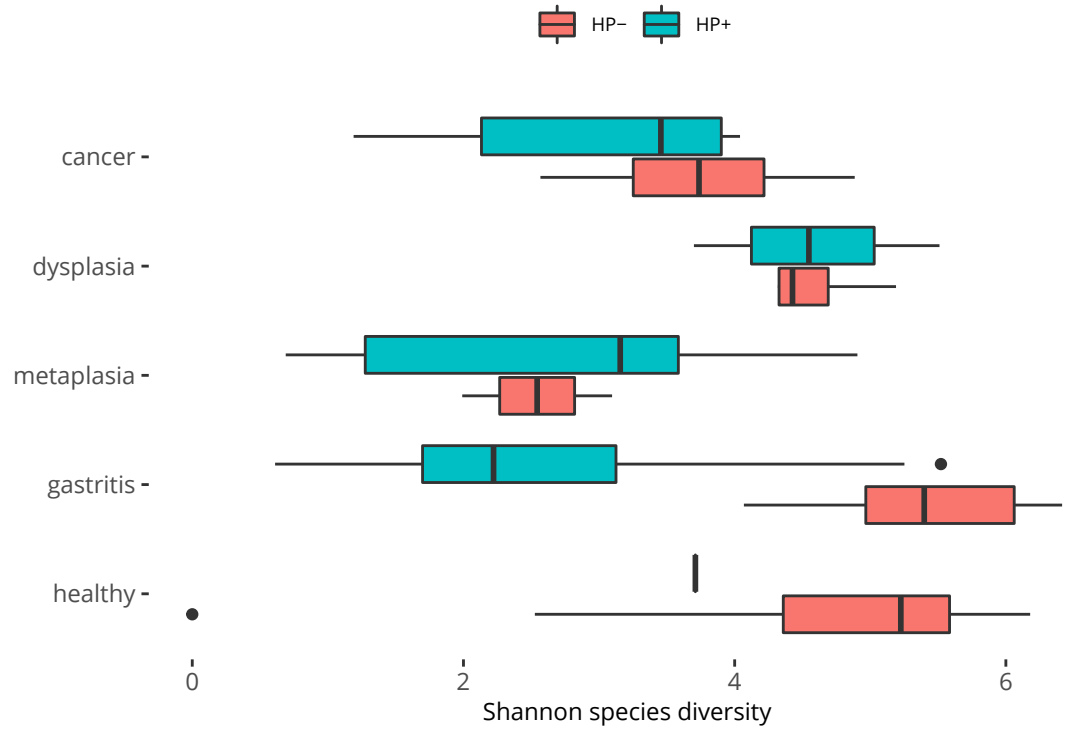

Figure S8: Shannon species diversity and disease progress (ERP023334). Helicobacter pylori positive (Hp+) and negative (Hp-) samples are distinguished.

We performed supervised learning of disease progress status with random forests on two thirds of the combined data set, with evaluation on the remaining third. Relative abundances summarized at the species level were used as the analysis substrate. Table S6 provides the classification results.

Table S6: Classification results on the disease stage evaluation subset, data set SRP070925. Predictions are in columns. Multiclass AUC:0.936.

stage	healthy	gastritis	meta-plasia	early cancer	adv. cancer
healthy	22				
gastritis		10			
metaplasia		4	2	3	
early cancer				7	1
advanced cancer				5	7

disease location

Using unweighted UniFrac distance on ASVs (amplicon sequence variants) we obtain better MDS separation of normal/peripherical/tumor samples than reported in [22], using the same dataset, whether without (not shown) or with addition of samples from healthy donors, figure S9.

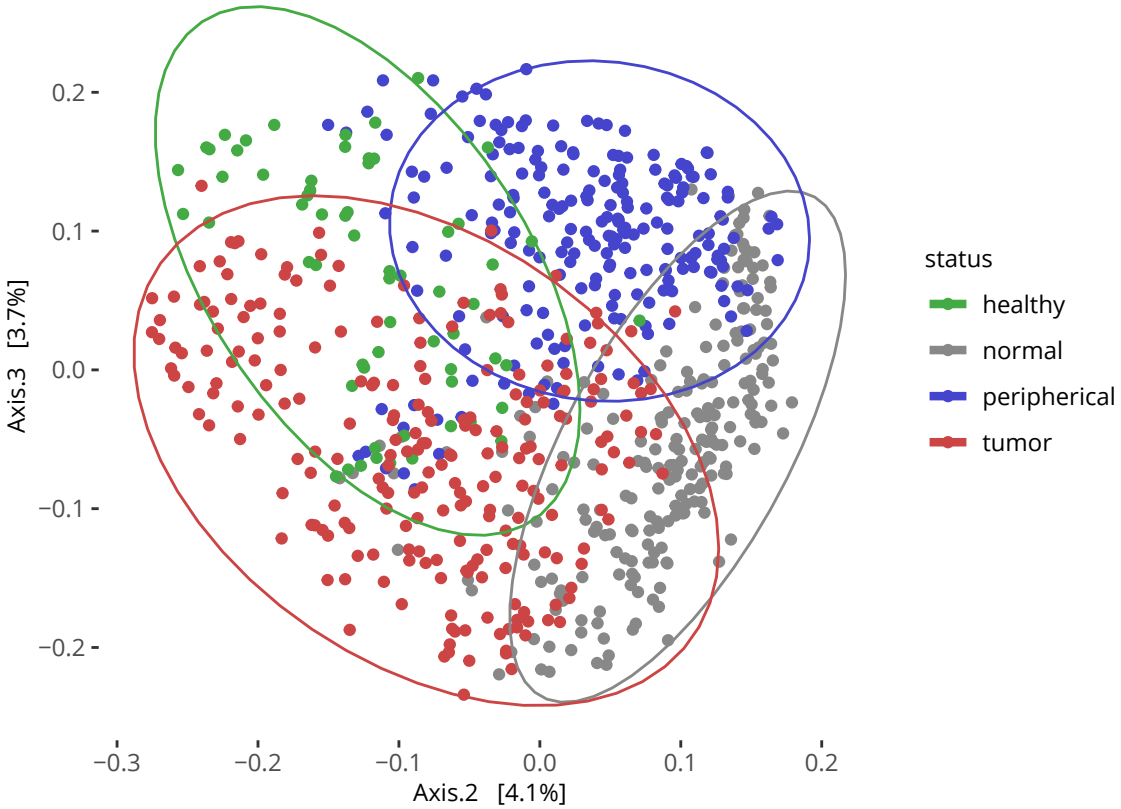

Figure S9: Multi-dimensional scaling of the disease status dataset SRP128749. Unweighted UniFrac of ASVs is used as the distance metric.

We performed two supervised learning experiments on the combined data set, one with a two-thirds training, one-third evaluation setup and a second using one additional data set SRP172818 (n=173) also containing triplets as the cross-validation set. All three data sets are from Chinese cohorts and have been analysed using the 16S variable regions V3-V4 combined on the Illumina MiSeq. Table S7 provides the classification results on the combined SRP128749 and SRP200169 data set. Table S8 provides the classification results on the cross-validation data set SRP172818.

Table S7: Combined SRP128749 and SRP200169 evaluation results. Predictions are in columns. Multiclass AUC:0.842

status	healthy	normal	peripherical	tumor
healthy	22		2	4
normal	1	37	20	11
peripherical	3	10	35	20
tumor		11	22	47

Table S8: SRP172818 cross-validation results. Predictions are in columns. Multiclass AUC:0.906

status	healthy	normal	peripherical	tumor
healthy				
normal		45	8	4
peripherical		7	41	9
tumor		4	7	48

relevant species in GC

We dispose of four datasets with the metadata required for the association of species with tumor status, whether from a disease progress or tumor/normal status standpoint. We choose to process datasets individually because of possible regional differences and retrieve the top 50 differentiating species from the random forest models, which we train on the datasets as a whole, so as to maximize performance. We provide sequence counts of these top 50 species to Spiec Easi for ecological network generation. We retain only connected nodes for display. Figure 2 in the main manuscript provides the result for the two tumor/peripherical/normal datasets SRP128749 and SRP172818 alongside for comparison. Figure S10 below provides the same for the disease progress data set SRP070925, Fig. S11 for the disease progress data set ERP023334.

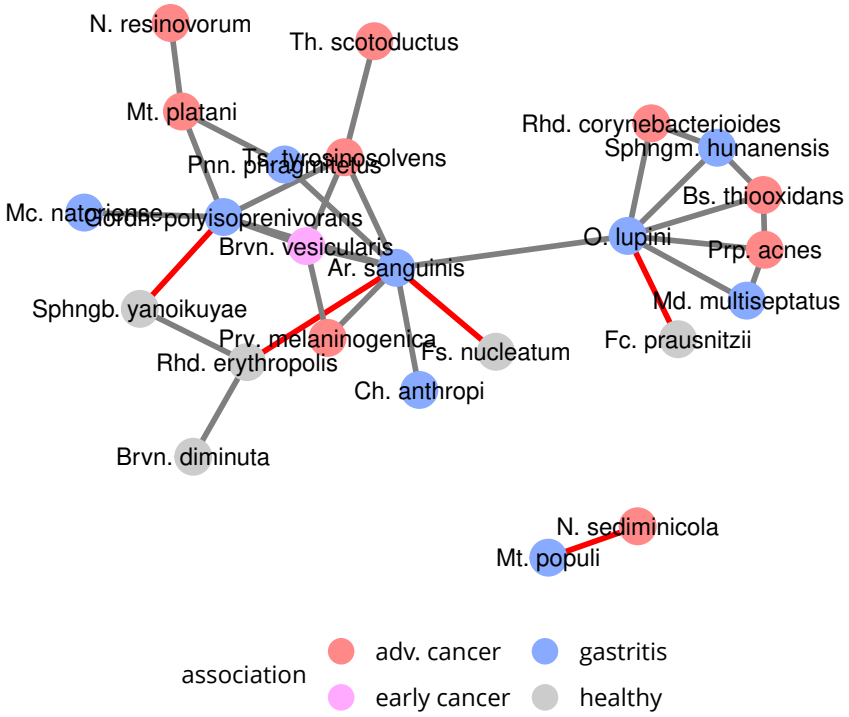

Figure S10: Interaction network between species relevant for disease progress (SRP070925). The top 50 species relevant for distinction between healthy the three disease stages are displayed. Only species with interactions are shown. Co-exclusion interactions are displayed in red.

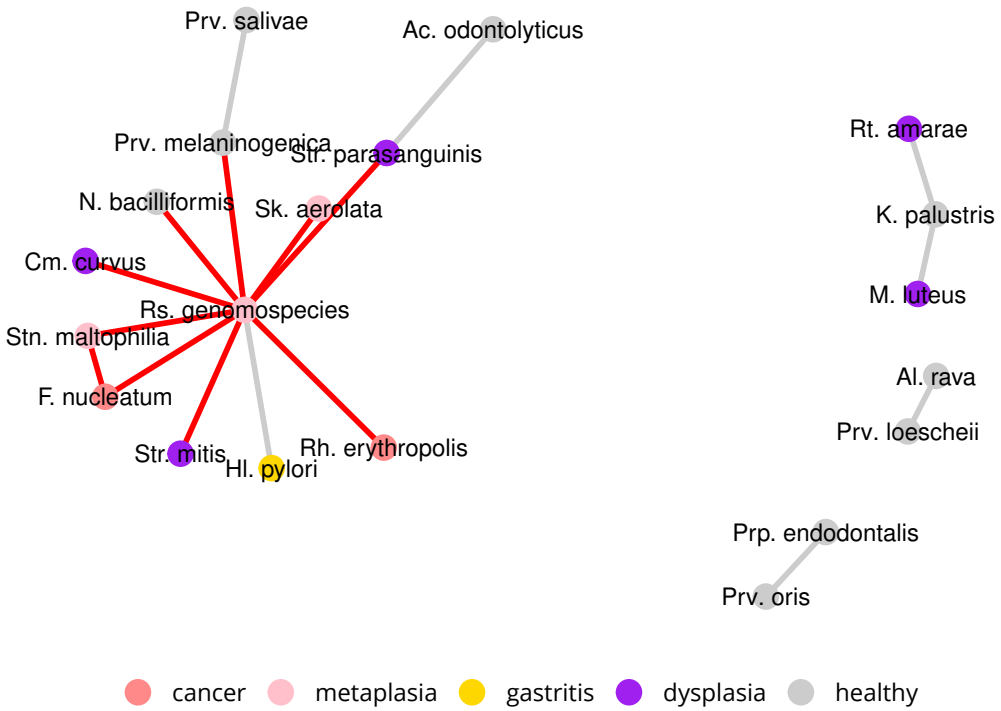

Figure S11: Interaction network between species relevant for disease progress (ERP023334). The top 50 species relevant for distinction between healthy four disease stages are displayed. Only species with interactions are shown. Co-exclusion interactions are displayed in red.

We can further investigate species differences by inferring prevalence differences between disease states of samples, using χ^2 testing, tables S4-S8. P-values were computed with Monte Carlo tests, 10,000 replicates.

Table S9: Prevalence differences between sample locations, SRP172818. Pearson's χ^2 p-values were computed by Monte Carlo simulation.

species	association	pvalue		normal	peripherical	tumor	count
Fusobacterium mortiferum	normal	1.0e-03	***	9/57 (15.8%)	0/57 (0.0%)	0/59 (0.0%)	9
Streptomyces atroolivaceus	normal	3.0e-03	**	7/57 (12.3%)	1/57 (1.8%)	0/59 (0.0%)	8
Peptostreptococcus stomatis	peripherical,tumor	1.3e-02	*	15/57 (26.3%)	24/57 (42.1%)	32/59 (54.2%)	71
Corynebacterium tuberculostearicum	tumor	1.0e-03	***	2/57 (3.5%)	1/57 (1.8%)	13/59 (22.0%)	16
Rudaeicoccus suwonensis	tumor	1.0e-03	***	3/57 (5.3%)	7/57 (12.3%)	18/59 (30.5%)	28
Propionibacterium acnes	tumor	1.0e-03	***	25/57 (43.9%)	22/57 (38.6%)	43/59 (72.9%)	90
Bifidobacterium dentium	tumor	1.0e-03	***	2/57 (3.5%)	3/57 (5.3%)	14/59 (23.7%)	19
Actinomyces odontolyticus	tumor	3.0e-03	**	1/57 (1.8%)	7/57 (12.3%)	14/59 (23.7%)	22
Deinococcus citri	tumor	3.0e-03	**	1/57 (1.8%)	0/57 (0.0%)	8/59 (13.6%)	9
Fusobacterium periodonticum	tumor	4.0e-03	**	1/57 (1.8%)	6/57 (10.5%)	12/59 (20.3%)	19
Rothia mucilaginosa	tumor	5.0e-03	**	10/57 (17.5%)	11/57 (19.3%)	25/59 (42.4%)	46
Campylobacter rectus	tumor	5.0e-03	**	1/57 (1.8%)	0/57 (0.0%)	8/59 (13.6%)	9
Fusobacterium nucleatum	tumor	6.0e-03	**	8/57 (14.0%)	12/57 (21.1%)	23/59 (39.0%)	43
Helicobacter pylori	tumor	3.7e-02	*	47/57 (82.5%)	51/57 (89.5%)	57/59 (96.6%)	155
Prevotella melaninogenica	tumor	4.2e-02	*	2/57 (3.5%)	4/57 (7.0%)	10/59 (16.9%)	16
Parvimonas micra	tumor	7.7e-02		15/57 (26.3%)	14/57 (24.6%)	25/59 (42.4%)	54

Table S10: Prevalence differences between sample locations, SRP128749. Pearson's χ^2 p-values were computed by Monte Carlo simulation.

species	association	pvalue		normal	peripherical	tumor	count
Nocardioides szechwanensis	normal	1.0e-03	***	9/225 (4.0%)	1/215 (0.5%)	0/229 (0.0%)	10
Bifidobacterium longum	normal	1.0e-03	***	42/225 (18.7%)	4/215 (1.9%)	18/229 (7.9%)	64
Prevotella stercorea	normal	1.0e-03	***	22/225 (9.8%)	2/215 (0.9%)	9/229 (3.9%)	33
Clostridium cellulovorans	normal	1.0e-03	***	16/225 (7.1%)	2/215 (0.9%)	1/229 (0.4%)	19
Roseburia inulinivorans	normal	1.0e-03	***	10/225 (4.4%)	0/215 (0.0%)	2/229 (0.9%)	12
Fusobacterium mortiferum	normal	1.0e-03	***	25/225 (11.1%)	3/215 (1.4%)	1/229 (0.4%)	29
Bacteroides uniformis	normal	3.0e-03	**	17/225 (7.6%)	2/215 (0.9%)	9/229 (3.9%)	28
Barnesiella intestinihominis	normal	3.0e-03	**	7/225 (3.1%)	0/215 (0.0%)	0/229 (0.0%)	7
Deinococcus aetherius	normal	3.0e-03	**	6/225 (2.7%)	0/215 (0.0%)	0/229 (0.0%)	6
Sulfurospirillum deleyianum	normal	5.0e-03	**	6/225 (2.7%)	0/215 (0.0%)	0/229 (0.0%)	6
Nitrospira japonica	normal	6.0e-03	**	11/225 (4.9%)	2/215 (0.9%)	2/229 (0.9%)	15
Parabacteroides merdae	normal	7.0e-03	**	5/225 (2.2%)	0/215 (0.0%)	0/229 (0.0%)	5
Faecalibacterium prausnitzii	normal,tumor	1.0e-03	***	127/225 (56.4%)	71/215 (33.0%)	111/229 (48.5%)	309
Ruminococcus bromii	normal,tumor	1.0e-03	***	57/225 (25.3%)	14/215 (6.5%)	42/229 (18.3%)	113
Arthrobacter oxydans	normal,tumor	2.0e-03	**	47/225 (20.9%)	28/215 (13.0%)	62/229 (27.1%)	137
Pyramidobacter piscolens	normal,tumor	4.0e-03	**	17/225 (7.6%)	2/215 (0.9%)	11/229 (4.8%)	30
Atopobium rimae	normal,tumor	6.0e-03	**	13/225 (5.8%)	4/215 (1.9%)	19/229 (8.3%)	36
Roseomonas gilardii	peripherical	3.0e-03	**	4/225 (1.8%)	20/215 (9.3%)	12/229 (5.2%)	36
Sphingomonas yabuuchiae	peripherical	4.0e-03	**	66/225 (29.3%)	94/215 (43.7%)	82/229 (35.8%)	242
Helicobacter pylori	peripherical,tumor	5.7e-02		155/225 (68.9%)	169/215 (78.6%)	175/229 (76.4%)	499
Corynebacterium tuberculostearicum	tumor	1.0e-03	***	10/225 (4.4%)	8/215 (3.7%)	30/229 (13.1%)	48
Propionibacterium acnes	tumor	1.0e-03	***	82/225 (36.4%)	65/215 (30.2%)	140/229 (61.1%)	287
Gardnerella vaginalis	tumor	1.0e-03	***	2/225 (0.9%)	1/215 (0.5%)	12/229 (5.2%)	15
Thermus scotoductus	tumor	1.0e-03	***	63/225 (28.0%)	48/215 (22.3%)	93/229 (40.6%)	204
Parvimonas micra	tumor	1.0e-03	***	39/225 (17.3%)	40/215 (18.6%)	85/229 (37.1%)	164
Catonella morbi	tumor	1.0e-03	***	7/225 (3.1%)	8/215 (3.7%)	29/229 (12.7%)	44
Peptostreptococcus stomatis	tumor	1.0e-03	***	52/225 (23.1%)	60/215 (27.9%)	130/229 (56.8%)	242
Fusobacterium nucleatum	tumor	1.0e-03	***	34/225 (15.1%)	45/215 (20.9%)	82/229 (35.8%)	161
Leptotrichia wadei	tumor	1.0e-03	***	13/225 (5.8%)	20/215 (9.3%)	40/229 (17.5%)	73
Sphingomonas faeni	tumor	1.0e-03	***	29/225 (12.9%)	47/215 (21.9%)	71/229 (31.0%)	147
Campylobacter showae	tumor	1.0e-03	***	0/225 (0.0%)	1/215 (0.5%)	14/229 (6.1%)	15
Corynebacterium mucifaciens	tumor	3.0e-03	**	4/225 (1.8%)	10/215 (4.7%)	21/229 (9.2%)	35
Filifactor alocis	tumor	3.0e-03	**	9/225 (4.0%)	13/215 (6.0%)	27/229 (11.8%)	49
Prevotella melaninogenica	tumor	5.3e-01		13/225 (5.8%)	14/215 (6.5%)	19/229 (8.3%)	46

Table S11: Prevalence differences between disease stages, SRP070925. Pearson's χ^2 p-values were computed by Monte Carlo simulation.

species	association	pvalue		advanced.cancer	early.cancer	gastritis	metaplasia	count
Peptostreptococcus stomatis	advanced cancer	4.0e-01		3/20 (15.0%)	0/20 (0.0%)	1/20 (5.0%)	1/20 (5.0%)	5
Novosphingobium sediminicola	early cancer,advanced cancer	1.0e-03	***	15/20 (75.0%)	17/20 (85.0%)	0/20 (0.0%)	6/20 (30.0%)	38
Methylobacterium populi	gastritis	1.0e-03	***	7/20 (35.0%)	6/20 (30.0%)	17/20 (85.0%)	8/20 (40.0%)	38
Sphingomonas hunanensis	gastritis	9.0e-03	**	1/20 (5.0%)	1/20 (5.0%)	8/20 (40.0%)	3/20 (15.0%)	13
Sphingobium amiense	gastritis	1.0e-02	**	0/20 (0.0%)	0/20 (0.0%)	4/20 (20.0%)	0/20 (0.0%)	4
Parvimonas micra	gastritis,advanced cancer	4.8e-01		3/20 (15.0%)	1/20 (5.0%)	2/20 (10.0%)	0/20 (0.0%)	6
Sphingomonas faeni	gastritis,early cancer	1.0e-02	**	2/20 (10.0%)	5/20 (25.0%)	10/20 (50.0%)	2/20 (10.0%)	19
Modestobacter multiseptatus	gastritis,metaplasia	1.0e-03	***	0/20 (0.0%)	1/20 (5.0%)	10/20 (50.0%)	7/20 (35.0%)	18
Hyphomonas polymorpha	gastritis,metaplasia	1.0e-03	***	4/20 (20.0%)	0/20 (0.0%)	16/20 (80.0%)	7/20 (35.0%)	27
Paenibacillus humicus	gastritis,metaplasia	4.0e-03	**	10/20 (50.0%)	12/20 (60.0%)	18/20 (90.0%)	18/20 (90.0%)	58
Prevotella melaninogenica	gastritis,metaplasia	4.8e-01		3/20 (15.0%)	4/20 (20.0%)	6/20 (30.0%)	7/20 (35.0%)	20
Helicobacter pylori	gastritis,metaplasia	6.3e-01		13/20 (65.0%)	12/20 (60.0%)	14/20 (70.0%)	16/20 (80.0%)	55
Fusobacterium nucleatum	metaplasia,advanced cancer	1.9e-01		8/20 (40.0%)	4/20 (20.0%)	2/20 (10.0%)	5/20 (25.0%)	19

Table S12: Prevalence differences between disease stages, ERP023334. Pearson's χ^2 p-values were computed by Monte Carlo simulation.

species	association	pvalue		cancer	dysplasia	gastritis	healthy	metaplasia	count
Corynebacterium pseudodiphtheriticum	dysplasia	1.0e-03	***	0/10 (0.0%)	4/8 (50.0%)	0/44 (0.0%)	1/22 (4.5%)	0/9 (0.0%)	5
Parvimonas micra	gastritis	7.8e-01		0/10 (0.0%)	0/8 (0.0%)	2/44 (4.5%)	0/22 (0.0%)	0/9 (0.0%)	2
Staphylococcus hominis	gastritis,dysplasia	3.0e-03	**	0/10 (0.0%)	5/8 (62.5%)	11/44 (25.0%)	3/22 (13.6%)	0/9 (0.0%)	19
Prevotella fusca	healthy	1.0e-03	***	0/10 (0.0%)	0/8 (0.0%)	0/44 (0.0%)	8/22 (36.4%)	0/9 (0.0%)	8
Tannerella forsythia	healthy	2.0e-03	**	3/10 (30.0%)	3/8 (37.5%)	13/44 (29.5%)	16/22 (72.7%)	1/9 (11.1%)	36
Prevotella loescheii	healthy	2.0e-03	**	3/10 (30.0%)	1/8 (12.5%)	9/44 (20.5%)	15/22 (68.2%)	1/9 (11.1%)	29
Prevotella oulorum	healthy	2.0e-03	**	3/10 (30.0%)	3/8 (37.5%)	12/44 (27.3%)	17/22 (77.3%)	2/9 (22.2%)	37
Prevotella veroralis	healthy	2.0e-03	**	1/10 (10.0%)	1/8 (12.5%)	6/44 (13.6%)	13/22 (59.1%)	1/9 (11.1%)	22
Treponema amylovorum	healthy	2.0e-03	**	0/10 (0.0%)	0/8 (0.0%)	3/44 (6.8%)	8/22 (36.4%)	0/9 (0.0%)	11
Prevotella dentalis	healthy	3.0e-03	**	1/10 (10.0%)	1/8 (12.5%)	2/44 (4.5%)	9/22 (40.9%)	0/9 (0.0%)	13
Prevotella pallens	healthy	3.0e-03	**	3/10 (30.0%)	3/8 (37.5%)	17/44 (38.6%)	17/22 (77.3%)	1/9 (11.1%)	41
Haemophilus sputorum	healthy	3.0e-03	**	0/10 (0.0%)	0/8 (0.0%)	0/44 (0.0%)	6/22 (27.3%)	0/9 (0.0%)	6
Treponema denticola	healthy	4.0e-03	**	2/10 (20.0%)	1/8 (12.5%)	6/44 (13.6%)	11/22 (50.0%)	0/9 (0.0%)	20
Propionibacterium acnes	healthy	4.5e-02	*	1/10 (10.0%)	2/8 (25.0%)	8/44 (18.2%)	11/22 (50.0%)	2/9 (22.2%)	24
Porphyromonas endodontalis	healthy,cancer	1.0e-03	***	4/10 (40.0%)	2/8 (25.0%)	9/44 (20.5%)	19/22 (86.4%)	3/9 (33.3%)	37
Alloprevotella rava	healthy,cancer	1.0e-03	***	4/10 (40.0%)	2/8 (25.0%)	8/44 (18.2%)	17/22 (77.3%)	1/9 (11.1%)	32
Solobacterium moorei	healthy,cancer	1.0e-03	***	3/10 (30.0%)	2/8 (25.0%)	6/44 (13.6%)	13/22 (59.1%)	0/9 (0.0%)	24
Actinomyces graevenitzii	healthy,dysplasia	1.0e-03	***	1/10 (10.0%)	5/8 (62.5%)	12/44 (27.3%)	15/22 (68.2%)	2/9 (22.2%)	35
Actinomyces odontolyticus	healthy,dysplasia	1.0e-03	***	5/10 (50.0%)	8/8 (100.0%)	18/44 (40.9%)	19/22 (86.4%)	4/9 (44.4%)	54
Prevotella oris	healthy,dysplasia	1.0e-03	***	2/10 (20.0%)	5/8 (62.5%)	12/44 (27.3%)	18/22 (81.8%)	2/9 (22.2%)	39
Capnocytophaga gingivalis	healthy,dysplasia	1.0e-03	***	2/10 (20.0%)	4/8 (50.0%)	8/44 (18.2%)	15/22 (68.2%)	2/9 (22.2%)	31
Selenomonas dianae	healthy,dysplasia	1.0e-03	***	3/10 (30.0%)	4/8 (50.0%)	12/44 (27.3%)	18/22 (81.8%)	1/9 (11.1%)	38
Lautropia mirabilis	healthy,dysplasia	1.0e-03	***	2/10 (20.0%)	5/8 (62.5%)	10/44 (22.7%)	16/22 (72.7%)	1/9 (11.1%)	34
Neisseria elongata	healthy,dysplasia	1.0e-03	***	1/10 (10.0%)	3/8 (37.5%)	9/44 (20.5%)	17/22 (77.3%)	1/9 (11.1%)	31
Campylobacter curvus	healthy,dysplasia	1.0e-03	***	4/10 (40.0%)	6/8 (75.0%)	17/44 (38.6%)	19/22 (86.4%)	3/9 (33.3%)	49
Aggregatibacter segnis	healthy,dysplasia	1.0e-03	***	0/10 (0.0%)	2/8 (25.0%)	7/44 (15.9%)	12/22 (54.5%)	1/9 (11.1%)	22
Haemophilus parainfluenzae	healthy,dysplasia	1.0e-03	***	4/10 (40.0%)	8/8 (100.0%)	17/44 (38.6%)	19/22 (86.4%)	2/9 (22.2%)	50
Porphyromonas catoniae	healthy,dysplasia	2.0e-03	**	2/10 (20.0%)	5/8 (62.5%)	16/44 (36.4%)	17/22 (77.3%)	2/9 (22.2%)	42
Alloprevotella tannerae	healthy,dysplasia	2.0e-03	**	2/10 (20.0%)	4/8 (50.0%)	16/44 (36.4%)	17/22 (77.3%)	1/9 (11.1%)	40
Veillonella atypica	healthy,dysplasia	2.0e-03	**	4/10 (40.0%)	7/8 (87.5%)	16/44 (36.4%)	15/22 (68.2%)	1/9 (11.1%)	43
Veillonella parvula	healthy,dysplasia	3.0e-03	**	3/10 (30.0%)	6/8 (75.0%)	15/44 (34.1%)	15/22 (68.2%)	1/9 (11.1%)	40
Prevotella intermedia	healthy,dysplasia	4.0e-03	**	1/10 (10.0%)	2/8 (25.0%)	6/44 (13.6%)	12/22 (54.5%)	0/9 (0.0%)	21
Prevotella salivae	healthy,dysplasia	4.0e-03	**	5/10 (50.0%)	7/8 (87.5%)	21/44 (47.7%)	19/22 (86.4%)	2/9 (22.2%)	54
Bradyrhizobium elkanii	healthy,dysplasia	4.0e-03	**	0/10 (0.0%)	3/8 (37.5%)	0/44 (0.0%)	3/22 (13.6%)	0/9 (0.0%)	6
Stenotrophomonas maltophilia	healthy,dysplasia	5.0e-03	**	0/10 (0.0%)	3/8 (37.5%)	6/44 (13.6%)	10/22 (45.5%)	1/9 (11.1%)	20
Streptococcus parasanguinis	healthy,dysplasia,cancer	1.0e-03	***	7/10 (70.0%)	8/8 (100.0%)	16/44 (36.4%)	20/22 (90.9%)	4/9 (44.4%)	55
Neisseria bacilliformis	healthy,dysplasia,cancer	1.0e-03	***	2/10 (20.0%)	2/8 (25.0%)	2/44 (4.5%)	11/22 (50.0%)	0/9 (0.0%)	17
Atopobium parvulum	healthy,dysplasia,cancer	3.0e-03	**	3/10 (30.0%)	3/8 (37.5%)	7/44 (15.9%)	12/22 (54.5%)	0/9 (0.0%)	25
Fusobacterium nucleatum	healthy,dysplasia,cancer	1.0e-02	**	8/10 (80.0%)	7/8 (87.5%)	25/44 (56.8%)	18/22 (81.8%)	2/9 (22.2%)	60

Table S13: Prevalence differences between disease stages, ERP023334. Pearson's χ^2 p-values were computed by Monte Carlo simulation.

species	association	pvalue		functional.dyspepsia	gastric.cancer	gastric.ulcer	count
Helicobacter pylori	functional dyspepsia	1.0e+00		4/6 (66.7%)	9/15 (60.0%)	8/13 (61.5%)	21
Methylobacterium radiotolerans	functional dyspepsia,gastric ulcer	1.0e-03	***	6/6 (100.0%)	5/15 (33.3%)	13/13 (100.0%)	24
Lactococcus lactis	gastric cancer	1.0e-03	***	2/6 (33.3%)	12/15 (80.0%)	1/13 (7.7%)	15
Peptostreptococcus stomatis	gastric cancer	3.6e-01		0/6 (0.0%)	2/15 (13.3%)	0/13 (0.0%)	2
Parvimonas micra	gastric cancer	1.0e+00		0/6 (0.0%)	1/15 (6.7%)	0/13 (0.0%)	1
Prevotella melaninogenica	gastric cancer,gastric ulcer	2.3e-01		0/6 (0.0%)	6/15 (40.0%)	4/13 (30.8%)	10
Fusobacterium nucleatum	gastric cancer,gastric ulcer	8.1e-01		1/6 (16.7%)	5/15 (33.3%)	5/13 (38.5%)	11

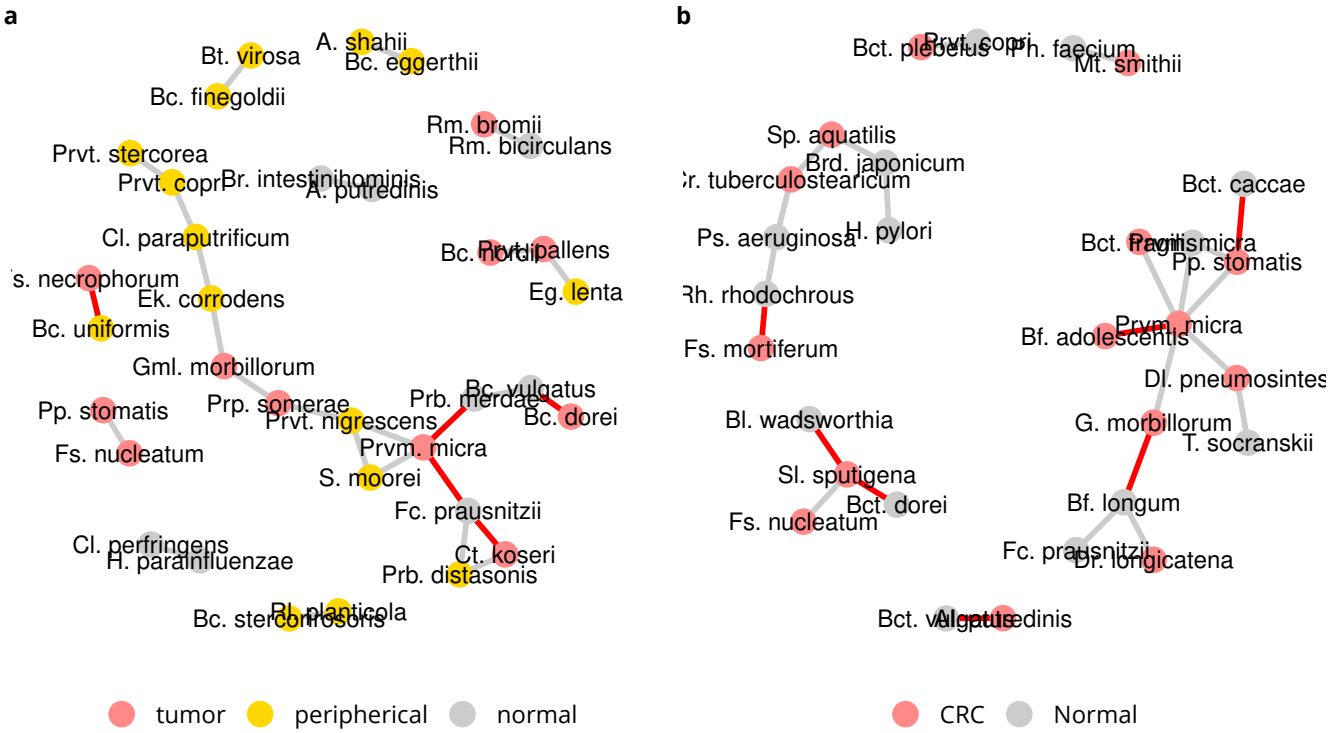

Figure S12: Discriminating species in CRC. Data sets a) SRP137015 and b) SRP076561. Only species with interactions are displayed. Location associations are based on maximum mean relative abundance. Co-exclusion is indicated in red.

comparison with CRC

We test two CRC data sets for presence and interactions of *F. nucleatum*, *P. micra* and *P. stomatis*. Data set SRP117763 (n=34, tumor-only) was published by [23] and data set SRP137015 (n=211, tumor/peripherical/normal) by [24, 25]. We find *F. nucleatum* in interaction with *P. stomatis* in SRP137015 and *P. micra* in interaction with *P. stomatis* in SRP117763, figure S12. Prevalence of *F. nucleatum* is over 70% in tumor samples in SRP117763, table S9 and at 48% in SRP137015, table S10.

Table S14: Prevalence differences between CRC subtypes, SRP117763. Pearson's χ^2 p-values were computed by Monte Carlo simulation.

species	association	pvalue		CMS1	CMS2	CMS3	count
Clostridium cadaveris	CMS1	8.0e-03	**	4/6 (66.7%)	2/13 (15.4%)	0/10 (0.0%)	6
Parvimonas micra	CMS1,CMS2	5.0e-02	*	3/6 (50.0%)	8/13 (61.5%)	1/10 (10.0%)	12
Peptostreptococcus stomatis	CMS1,CMS2	1.9e-01		2/6 (33.3%)	6/13 (46.2%)	1/10 (10.0%)	9
Prevotella melaninogenica	CMS1,CMS2	6.8e-01		1/6 (16.7%)	1/13 (7.7%)	0/10 (0.0%)	2
Fusobacterium nucleatum	CMS1,CMS2	1.0e+00		5/6 (83.3%)	10/13 (76.9%)	7/10 (70.0%)	22

Table S15: Prevalence differences between CRC sample locations, SRP137015. Pearson's χ^2 p-values were computed by Monte Carlo simulation.

species	association	pvalue		normal	peripherical	tumor	count
Prevotella melaninogenica	normal	1.0e+00		1/103 (1.0%)	0/46 (0.0%)	0/62 (0.0%)	1
Bacteroides vulgatus	normal,peripherical	3.0e-03	**	80/103 (77.7%)	38/46 (82.6%)	34/62 (54.8%)	152
Peptostreptococcus stomatis	peripherical,tumor	1.7e-01		12/103 (11.7%)	8/46 (17.4%)	14/62 (22.6%)	34
Campylobacter gracilis	tumor	1.0e-03	***	1/103 (1.0%)	1/46 (2.2%)	8/62 (12.9%)	10
Fusobacterium nucleatum	tumor	2.0e-03	**	20/103 (19.4%)	12/46 (26.1%)	30/62 (48.4%)	62
Parvimonas micra	tumor	3.0e-03	**	5/103 (4.9%)	5/46 (10.9%)	14/62 (22.6%)	24

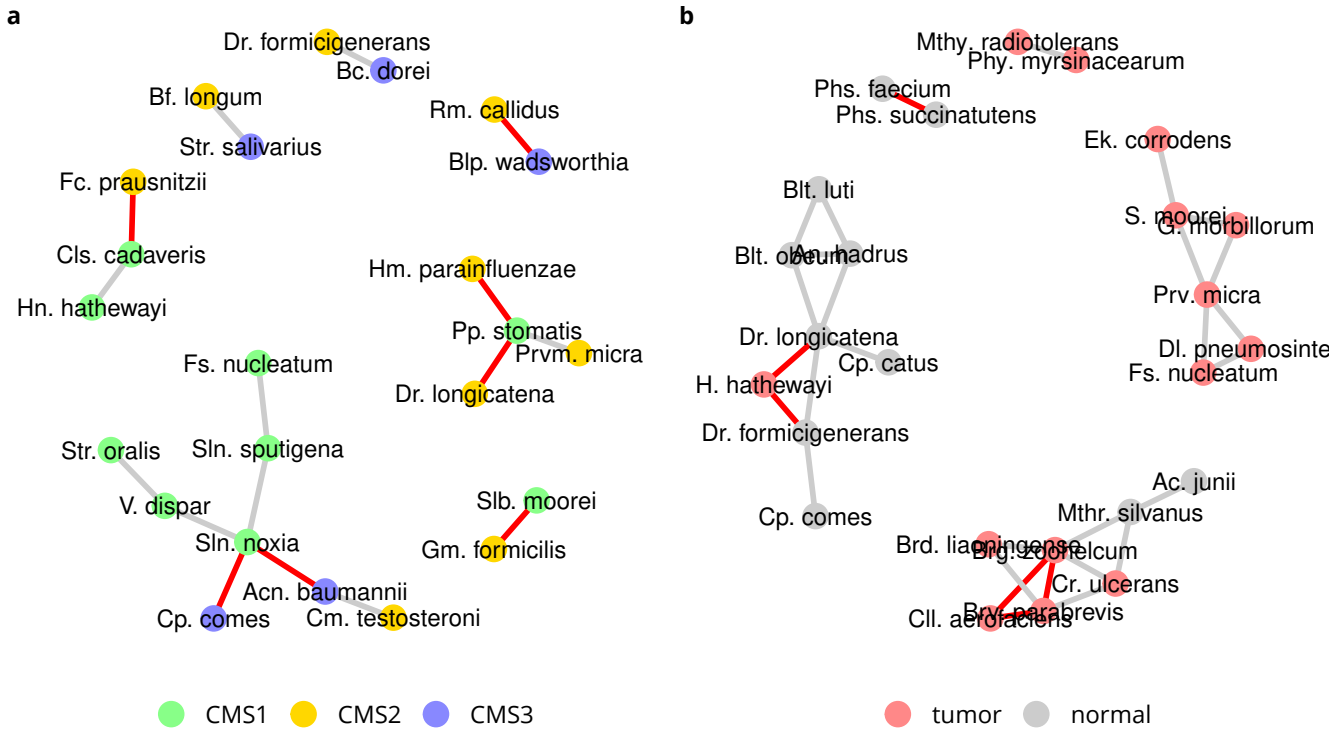

Figure S13: Discriminating species in CRC. Data sets a) SRP117763 and b) ERP005534 Only species with interactions are displayed. Location associations are based on maximum mean relative abundance. Co-exclusion is indicated in red.

Table S16: Prevalence differences between CRC sample locations, SRP076561. Pearson's χ^2 p-values were computed by Monte Carlo simulation.

species	association	pvalue	CRC	Normal	count
Fusobacterium nucleatum	tumor	0.13	19/26 (73.1%)	12/24 (50.0%)	31
Prevotella melaninogenica	tumor	1.00	1/26 (3.8%)	0/24 (0.0%)	1
Propionibacterium acnes	normal	0.15	8/26 (30.8%)	13/24 (54.2%)	21
Helicobacter pylori	normal	0.58	14/26 (53.8%)	15/24 (62.5%)	29
Parvimonas micra	normal	0.60	15/26 (57.7%)	16/24 (66.7%)	31
Peptostreptococcus stomatis	normal	1.00	16/26 (61.5%)	15/24 (62.5%)	31

Table S17: Prevalence differences between CRC sample locations, ERP005534. Pearson's χ^2 p-values were computed by Monte Carlo simulation.

species	association	pvalue	normal	tumor	count
Parvimonas micra		1.00	33/48 (68.8%)	33/48 (68.8%)	66
Prevotella melaninogenica	normal	0.51	2/48 (4.2%)	0/48 (0.0%)	2
Fusobacterium nucleatum	tumor	0.11	31/48 (64.6%)	39/48 (81.2%)	70
Peptostreptococcus stomatis	tumor	0.68	22/48 (45.8%)	25/48 (52.1%)	47

Bibliography

[1] Nielson T Baxter, Mack T Ruffin, Mary A M Rogers, and Patrick D Schloss. Microbiota-based model improves the sensitivity of fecal immunochemical test for detecting colonic lesions. *Genome Medicine*, 8(1):37, April 2016. doi: 10.1186/s13073-016-0290-3.

[2] Jacobo de la Cuesta-Zuluaga, Noel T Mueller, Vanessa Corrales-Agudelo, Eliana P Velásquez-Mejía, Jenny A Carmona, José M Abad, and Juan S Escobar. Metformin Is Associated With Higher Relative Abundance of Mucin-Degrading Akkermansia muciniphila and Several Short-Chain Fatty Acid-Producing Microbiota in the Gut. *Diabetes Care*, 40(1):54–62, January 2017. doi: 10.2337/dc16-1324.

[3] Mária Džunková, Andrés Moya, Jorge F Vázquez-Castellanos, Alejandro Artacho, Xinhua Chen, Ciaran Kelly, Giuseppe D'Auria, and Hideyuki Tamaki. Active and Secretory IgA-Coated Bacterial Fractions Elucidate Dysbiosis in Clostridium difficile Infection. *mSphere*, 1(3):e00101–16, June 2016. doi: 10.1128/mSphere.00101-16.

[4] Dirk Gevers, Subra Kugathasan, Lee A Denson, Yoshiki Vázquez-Baeza, Will Van Treuren, Boyu Ren, Emma Schwager, Dan Knights, Se Jin Song, Moran Yassour, Xochitl C Morgan, Aleksandar D Kostic, Chengwei Luo, Antonio González, Daniel McDonald, Yael Haberman, Thomas Walters, Susan Baker, Joel Rosh, Michael Stephens, Melvin Heyman, James Markowitz, Robert Baldassano, Anne Griffiths, Francisco Sylvester, David Mack, Sandra Kim, Wallace Crandall, Jeffrey Hyams, Curtis Huttenhower, Rob Knight, and Ramnik J Xavier. The Treatment-Naive Microbiome in New-Onset Crohn's Disease. *Cell host & microbe*, 15(3):382–392, March 2014. doi: 10.1016/j.chom.2014.02.005.

[5] James J Goedert, Gieira Jones, Xing Hua, Xia Xu, Guoqin Yu, Roberto Flores, Roni T Falk, Mitchell H Gail, Jianxin Shi, Jacques Ravel, and Heather Spencer Feigelson. Investigation of the association between the fecal microbiota and breast cancer in postmenopausal women: a population-based case-control pilot study. *Journal of the National Cancer Institute*, 107(8), August 2015. doi: 10.1093/jnci/djv147.

[6] Julia K Goodrich, Jillian L Waters, Angela C Poole, Jessica L Sutter, Omry Koren, Ran Blekhman, Michelle Beaumont, William Van Treuren, Rob Knight, Jordana T Bell, Timothy D Spector, Andrew G Clark, and Ruth E Ley. Human genetics shape the gut microbiome. *Cell*, 159(4):789–799, November 2014. doi: 10.1016/j.cell.2014.09.053.

[7] Andrea Hahn, Amit Sanyal, Geovanny F Perez, Anamaris M Colberg-Poley, Joseph Campos, Mary C Rose, and Marcos Pérez-Losada. Different next generation sequencing platforms produce different microbial profiles and diversity in cystic fibrosis sputum. *Journal of Microbiological Methods*, 130:95–99, November 2016. doi: 10.1016/j.mimet.2016.09.002.

[8] Deborah A Hogan, Sven D Willger, Emily L Dolben, Thomas H Hampton, Bruce A Stanton, Hilary G Morrison, Mitchell L Sogin, Julianna Czum, and Alix Ashare. Analysis of Lung Microbiota in Bronchoalveolar Lavage, Protected Brush and Sputum Samples from Subjects with Mild-To-Moderate Cystic Fibrosis Lung Disease. *PLoS ONE*, 11(3), 2016. doi: 10.1371/journal.pone.0149998.

[9] Emily J Johnson, Edith T Zemanick, Frank J Accurso, Brandie D Wagner, Charles E Robertson, and J Kirk Harris. Molecular Identification of Staphylococcus aureus in Airway Samples from Children with Cystic Fibrosis. *PLoS ONE*, 11(1), 2016. doi: 10.1371/journal.pone.0147643.t002.

[10] S Khanna, E Montassier, B Schmidt, R Patel, D Knights, D S Pardi, and P C Kashyap. Gut microbiome predictors of treatment response and recurrence in primary Clostridium difficileinfection. *Alimentary Pharmacology & Therapeutics*, 44(7):715–727, August 2016. doi: 10.1111/apt.13750.

[11] Yingying Lu, Jing Chen, Junyuan Zheng, Guoyong Hu, Jingjing Wang, Chunlan Huang, Lihong Lou, Xingpeng Wang, and Yue Zeng. Mucosal adherent bacterial dysbiosis in patients with colorectal adenomas. *Scientific Reports*, 6:26337, May 2016. doi: 10.1038/srep26337.

[12] Daniel McDonald, Embriette Hyde, Justine W Debelius, James T Morton, Antonio Gonzalez, Gail Ackermann, Alexander A Aksenov, Bahar Behsaz, Caitriona Brennan, Yingfeng Chen, Lindsay DeRight Goldasich, Pieter C Dorrestein, Robert R Dunn, Ashkaan K Fahimipour, James Gaffney, Jack A Gilbert, Grant Gogul, Jessica L Green, Philip Hugenholtz, Greg Humphrey, Curtis Huttenhower, Matthew A Jackson, Stefan Janssen, Dilip V Jeste, Lingjing Jiang, Scott T Kelley, Dan Knights, Tomasz Kosciolek, Joshua Ladau, Jeff Leach, Clarisse Marotz, Dmitry Meleshko, Alexey V Melnik, Jessica L Metcalf, Hosein Mohimani, Emmanuel Montassier, Jose Navas-Molina, Tanya T Nguyen, Shyamal Peddada, Pavel Pevzner, Katherine S Pollard, Gholamali Rahnavard, Adam Robbins-Pianka, Naseer Sangwan, Joshua Shorenstein, Larry Smarr, Se Jin Song, Timothy Spector, Austin D Swafford, Varykina G Thackray, Luke R Thompson, Anupriya Tripathi, Yoshiki Vázquez-Baeza, Alison Vrbanac, Paul Wischmeyer, Elaine Wolfe, Qiyun Zhu, American Gut Consortium , and Rob Knight. American Gut: an Open Platform for Citizen Science Microbiome Research. *mSystems*, 3(3), May 2018. doi: 10.1128/mSystems.00031-18.

[13] Jacquelyn S Meisel, Geoffrey D Hannigan, Amanda S Tyldsley, Adam J SanMiguel, Brendan P Hodkinson, Qi Zheng, and Elizabeth A Grice. Skin Microbiome Surveys Are Strongly Influenced by Experimental Design. *Journal of Investigative Dermatology*, 136(5):947–956, May 2016. doi: 10.1016/j.jid.2016.01.016.

[14] Elin Org, Yuna Blum, Silva Kasela, Margarete Mehrabian, Johanna Kuusisto, Antti J Kangas, Pasi Soininen, Zeneng Wang, Mika Ala-Korpela, Stanley L Hazen, Markku Laakso, and Aldons J Lusis. Relationships between gut microbiota, plasma metabolites, and metabolic syndrome traits in the METSIM cohort. *Genome Biology*, 18(1):70, April 2017. doi: 10.1186/s13059-017-1194-2.

[15] Alba Santiago, Marta Pozuelo, Maria Poca, Cristina Gely, Juan Camilo Nieto, Xavier Torras, Eva Román, David Campos, Guillaume Sarrabayrouse, Silvia Vidal, Edilmar Alvarado-Tapias, Francisco Guarner, German Soriano, Chaysavanh Manichanh, and Carlos Guarner. Alteration of the serum microbiome composition in cirrhotic patients with ascites. *Scientific Reports*, 6:25001, April 2016. doi: 10.1038/srep25001.

[16] Anna Maria Seekatz, Krishna Rao, Kavitha Santhosh, and Vincent Bensan Young. Dynamics of the fecal microbiome in patients with recurrent and nonrecurrent Clostridium difficile infection. *Genome Medicine*, 8(1): 47, April 2016. doi: 10.1186/s13073-016-0298-8.

[17] Kelly A Shaw, Madeline Bertha, Tatyana Hofmekler, Pankaj Chopra, Tommi Vatanen, Abhiram Srivatsa, Jarod Prince, Archana Kumar, Cary Sauer, Michael E Zwick, Glen A Satten, Aleksandar D Kostic, Jennifer G Mulle, Ramnik J Xavier, and Subra Kugathasan. Dysbiosis, inflammation, and response to treatment: a longitudinal study of pediatric subjects with newly diagnosed inflammatory bowel disease. *Genome Medicine*, 8(1):599, July 2016. doi: 10.1186/s13073-016-0331-y.

[18] Meera K Shenoy, Shoko Iwai, Din L Lin, William Worodria, Irene Ayakaka, Patrick Byanyima, Sylvia Kaswabuli, Serena Fong, Stephen Stone, Emily Chang, J Lucian Davis, Ali Ahmad Faruqi, Mark R Segal, Laurence Huang, and Susan V Lynch. Immune Response and Mortality Risk Relate to Distinct Lung Microbiomes in HIV-Pneumonia Patients. *American journal of respiratory and critical care medicine*, pages rccm.201603–0523OC, July 2016. doi: 10.1164/rccm.201603-0523OC.

[19] Hakdong Shin, Kenneth Price, Luong Albert, Jack Dodick, Lisa Park, and Maria Gloria Dominguez-Bello. Changes in the Eye Microbiota Associated with Contact Lens Wearing. *mBio*, 7(2):e00198, March 2016. doi: 10.1128/mBio.00198-16.

[20] Joshua S Son, Ling J Zheng, Leahana M Rowehl, Xinyu Tian, Yuanhao Zhang, Wei Zhu, Leighann Litcher-Kelly, Kenneth D Gadow, Grace Gathungu, Charles E Robertson, Diana Ir, Daniel N Frank, and Ellen Li. Comparison of Fecal Microbiota in Children with Autism Spectrum Disorders and Neurotypical Siblings in the Simons Simplex Collection. *PLoS ONE*, 10(10), 2015. doi: 10.1371/journal.pone.0137725.

[21] Tommi Vatanen, Aleksandar D Kostic, Eva d'Hennezel, Heli Siljander, Eric A Franzosa, Moran Yassour, Raivo Kolde, Hera Vlamakis, Timothy D Arthur, Anu-Maaria Hämäläinen, Aleksandr Peet, Vallo Tillmann, Raivo Uibo, Sergei Mokurov, Natalya Dorshakova, Jorma Ilonen, Suvi M Virtanen, Susanne J Szabo, Jeffrey A Porter, Harri Lähdesmäki, Curtis Huttenhower, Dirk Gevers, Thomas W Cullen, Mikael Knip, DIABIMMUNE Study Group, and Ramnik J Xavier. Variation in Microbiome LPS Immunogenicity Contributes to Autoimmunity in Humans. *Cell*, 165(4):842–853, May 2016. doi: 10.1016/j.cell.2016.04.007.

[22] Xiaosun Liu, Li Shao, Xia Liu, Feng Ji, Ying Mei, Yiwen Cheng, Fengping Liu, Chongxian Yan, Lanjuan Li, and Zongxin Ling. Alterations of gastric mucosal microbiota across different stomach microhabitats in a cohort of 276 patients with gastric cancer. *EBioMedicine*, 40:336–348, February 2019. doi: 10.1016/j.ebiom.2018.12.034.

[23] Rachel V Purcell, Martina Visnovska, Patrick J Biggs, Sebastian Schmeier, and Frank A Frizelle. Distinct gut microbiome patterns associate with consensus molecular subtypes of colorectal cancer. *Scientific Reports*, 7 (1):11590, September 2017. doi: 10.1038/s41598-017-11237-6.

[24] Vanessa L Hale, Patricio Jeraldo, Michael Mundy, Janet Yao, Gary Keeney, Nancy Scott, E Heidi Cheek, Jennifer Davidson, Megan Greene, Christine Martinez, John Lehman, Chandra Pettry, Erica Reed, Kelly Lyke, Bryan A White, Christian Diener, Osbaldo Resendis-Antonio, Jaime Gransee, Tumpa Dutta, Xuan-Mai Petterson, Lisa Boardman, David Larson, Heidi Nelson, and Nicholas Chia. Synthesis of multi-omic data and community metabolic models reveals insights into the role of hydrogen sulfide in colon cancer. *METHODS*, 149:59–68, October 2018. doi: 10.1016/j.ymeth.2018.04.024.

[25] Vanessa L Hale, Patricio Jeraldo, Jun Chen, Michael Mundy, Janet Yao, Sambhawa Priya, Gary Keeney, Kelly Lyke, Jason Ridlon, Bryan A White, Amy J French, Stephen N Thibodeau, Christian Diener, Osbaldo Resendis-Antonio, Jaime Gransee, Tumpa Dutta, Xuan-Mai Petterson, Jaeyun Sung, Ran Blekhman, Lisa Boardman, David Larson, Heidi Nelson, and Nicholas Chia. Distinct microbes, metabolites, and ecologies define the microbiome in deficient and proficient mismatch repair colorectal cancers. *Genome Medicine*, 10(1):78–13, October 2018. doi: 10.1186/s13073-018-0586-6.