

## **SUPPLEMENTAL METHODS**

### **Preprocessing and OTU selection:**

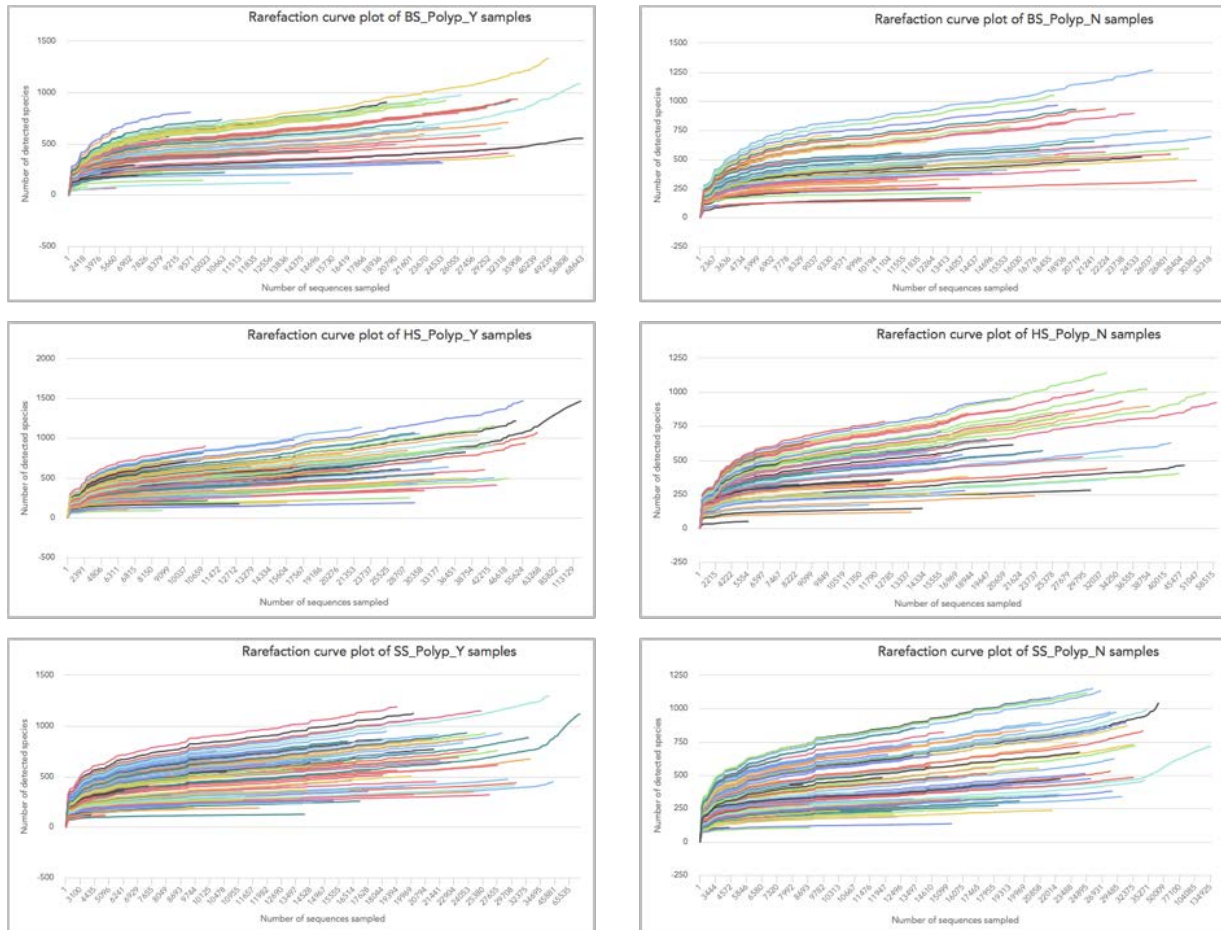
Sequences were preprocessed based on the following criteria: the sequences containing more than 5 ambiguous bases and sequences with more than 8 polymers were culled. All chimeric reads were found and removed using UCHIME (Edgar et al., 2011). The reference database used for chimera removal was RDP.gold (<https://rdp.cme.msu.edu>) (Release 11, Update 4; May 26, 2015). Undesirable lineages (Mitochondria-Chloroplast-Eukaryotes-Archaea) were removed using the Naïve Bayes classifier (Wang et al., 2007). Reads with the low-quality score were removed. Singletons (reads that appear just once) were discarded. Reads were trimmed at a fixed length. OTU selection was performed using UPARSE (Edgar, 2013). After finding OTUs or centroids, all the reads were mapped to OTUs using an identity threshold of 97%. An abundance table was constructed that included all the OTUs and their abundance for each sample. The OTUs were aligned to 16S reference sequence database GreenGenes (<http://greengenes.lbl.gov>) using the RDP classifier (Wang et al., 2007), and a taxonomic ID was generated for each OTU.

### **Beta diversity:**

Beta diversity was analyzed by generating a phylogenetic tree and this was followed by Unique Fraction (UniFrac) and principal coordinates analysis (PCoA). Specifically, the processed reads were aligned using UPGMA algorithm (Unweighted Pair Group Method with Arithmetic Mean) and a phylogenetic tree was produced using FastTree (<http://microbesonline.org/fasttree/>) (Price et al., 2009; Price et al., 2010) which makes an approximate-maximum-likelihood phylogenetic tree. Then weighted and unweighted UniFrac metrics were calculated using the tree and the results were visualized using principal coordinate analysis.

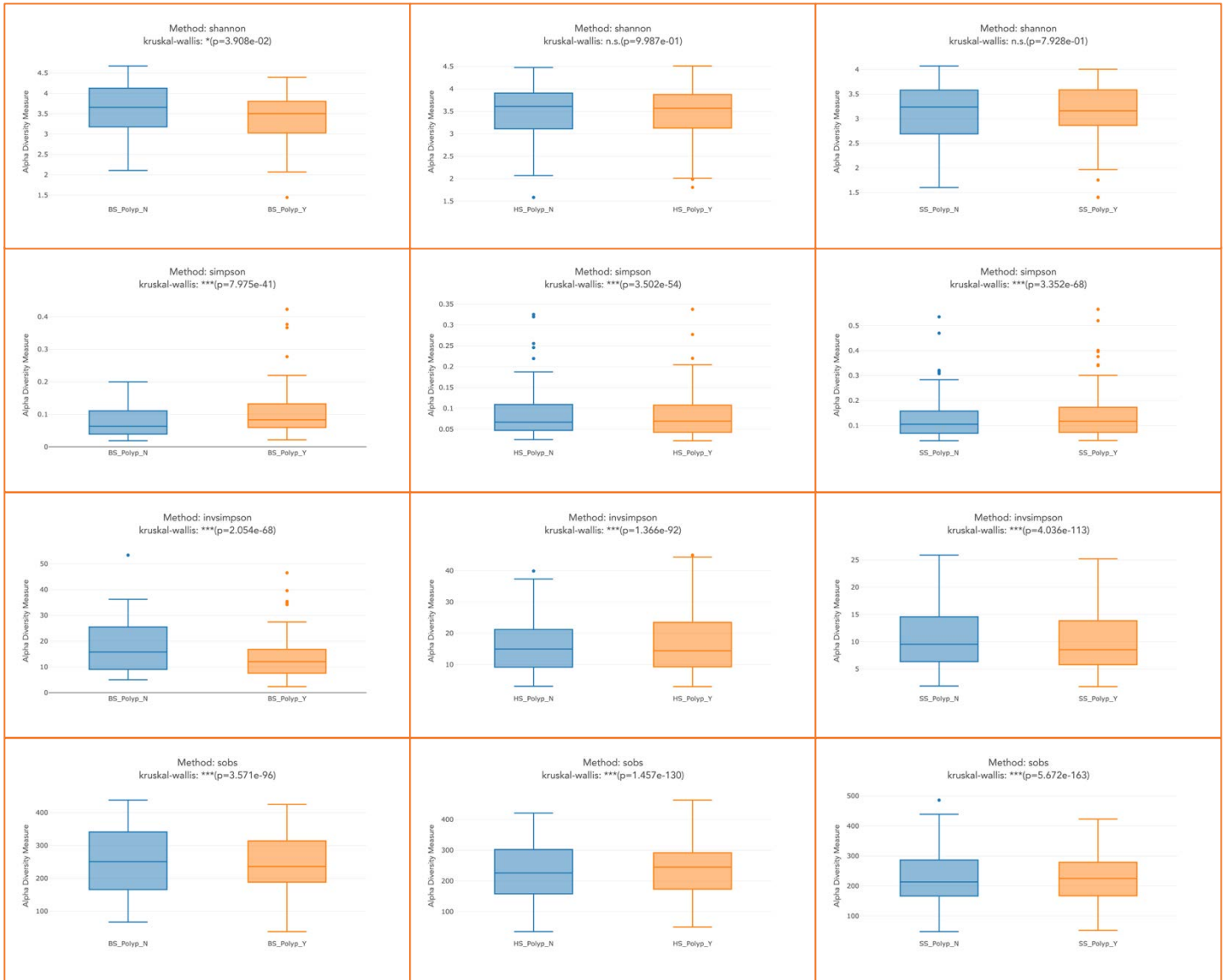
## SUPPLEMENTAL FIGURES

**Figure S1. Polyp dataset rarefaction plots for biopsy, stool, and rectal swab samples**



All plots have reached a plateau, which means sequencing depth is acceptable.

**Figure S2. Alpha diversity: results for biopsy, home stool, and rectal swab samples**



For the diversity indices of Shannon, Simpson, invSimpson, and observed species (sobs), the diversity was significantly different between polyp-Y and polyp-N groups except for shannon diversity index that was just significantly different in the biopsy dataset.

BS, biopsy samples; HS, home stool samples; n.s., non significant; Polyp-N, polyp-negative group; Polyp-Y, polyp-positive group; SS, rectal swab samples.

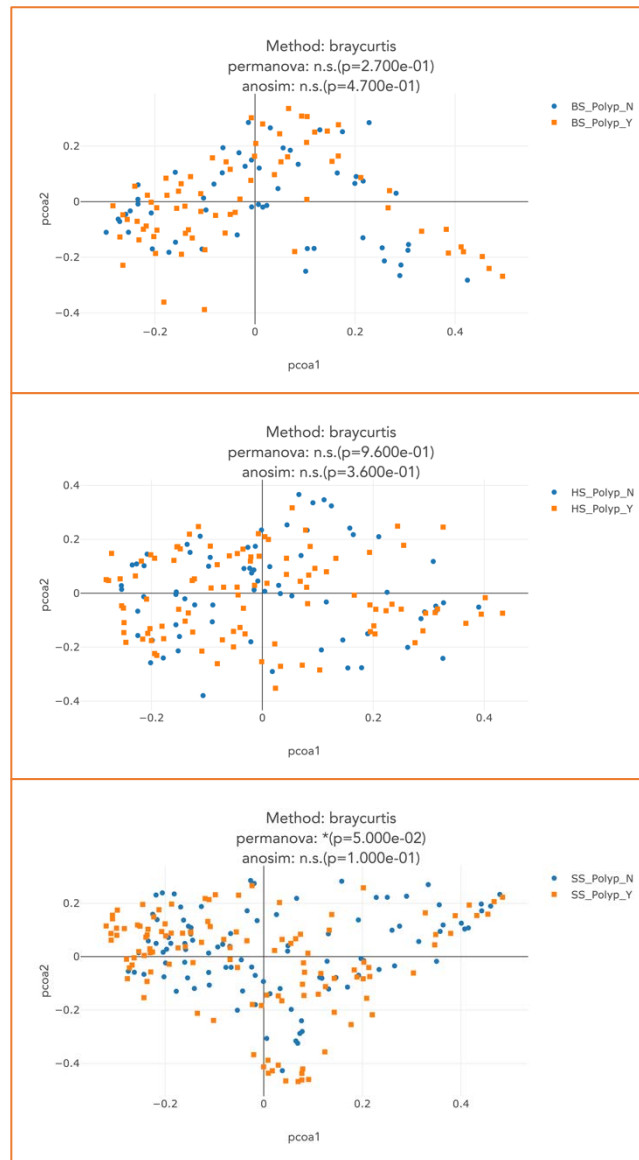
**Figure S3. Beta diversity: UniFrac PCoA plots of the biopsy, home stool, and rectal swab samples**



PERMANOVA test showed that weighted UniFrac distances in the biopsy group were significantly different between polyp-Y and polyp-N groups. ANOSIM couldn't find any significant difference between the polyp-Y and polyp-N groups in none of the datasets. The axes are principal components.

ANOSIM, analysis of similarities; BS, biopsy samples; HS, home stool samples; PERMANOVA, permutational multivariate analysis of variance; PCoA, principal coordinates analysis; Polyp-N, polyp-negative group; Polyp-Y, polyp-positive group; SS, rectal swab samples; UniFrac, unique fraction.

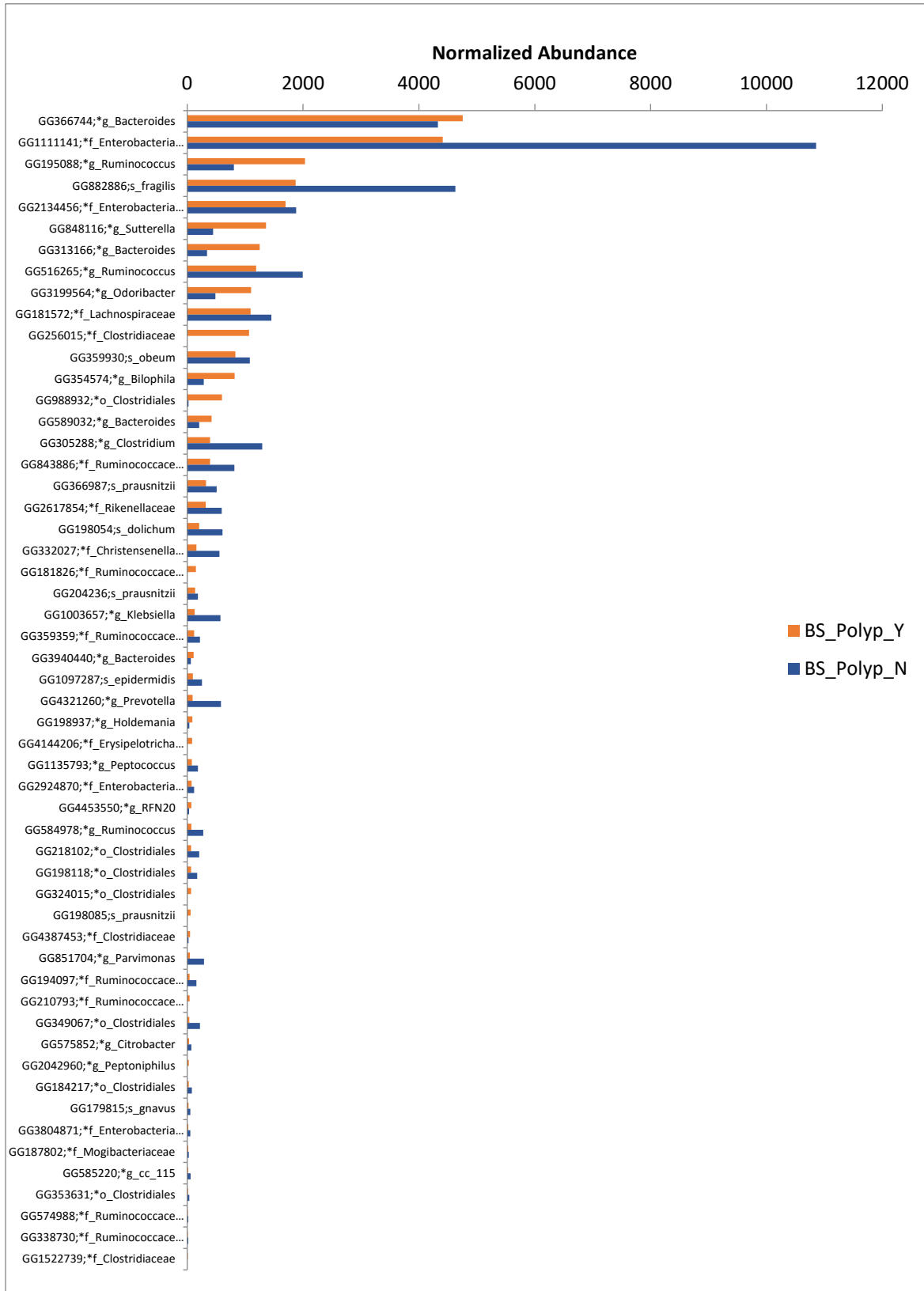
**Figure S4. Beta diversity: Bray-Curtis PCoA plots of the biopsy, home stool, and rectal swab samples**

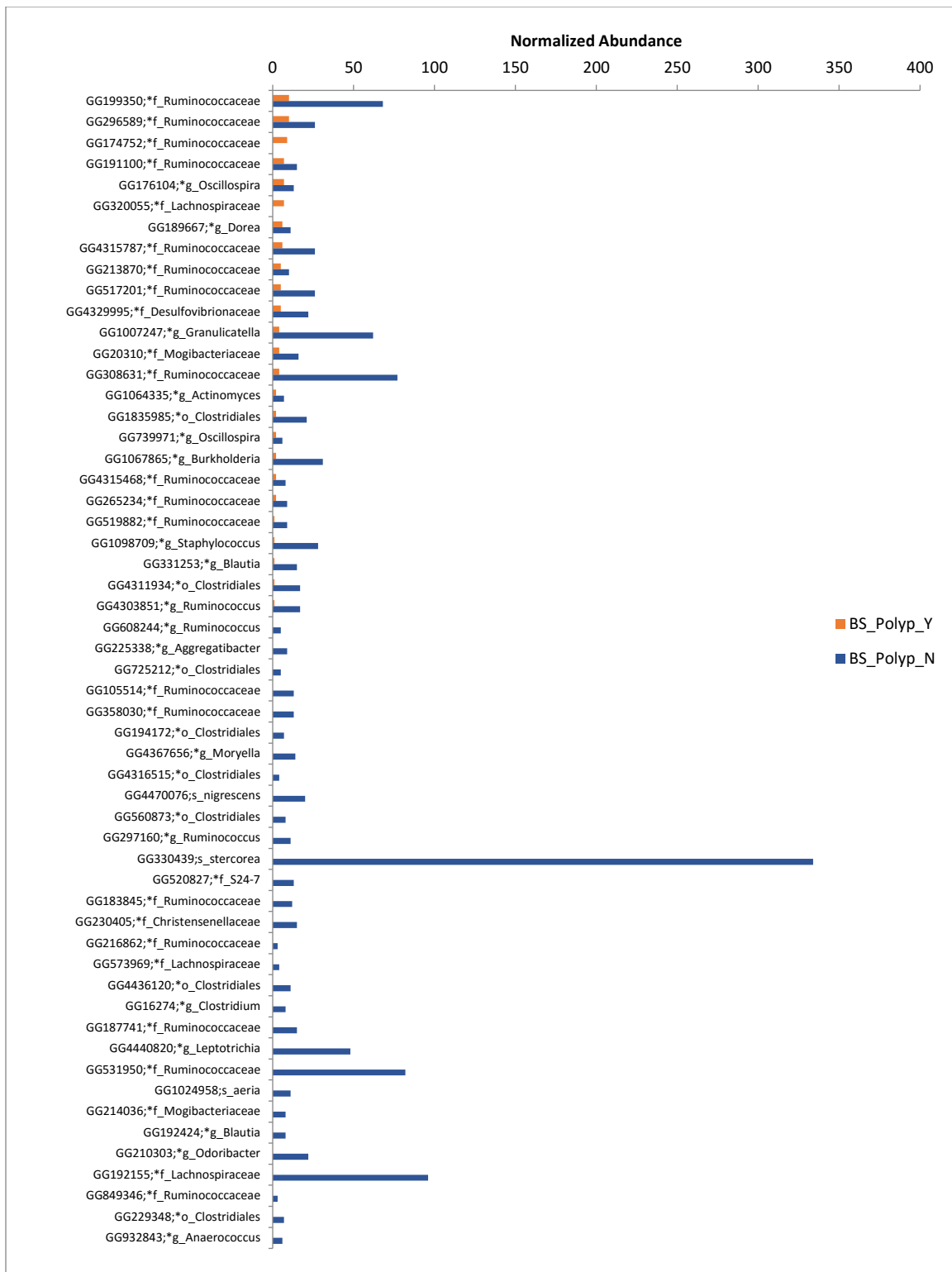


PERMANOVA test represented a significant difference between polyp-Y and polyp-N groups in the rectal swabs dataset. ANOSIM was not successful in finding a difference between polyp-Y and polyp-N groups in none of the datasets. The axes are principal components.

ANOSIM, analysis of similarities; BS, biopsy samples; HS, home stool samples; PERMANOVA, permutational multivariate analysis of variance; PCoA, principal coordinates analysis; Polyp-N, polyp-negative group; Polyp-Y, polyp-positive group; SS, rectal swab samples.

**Figure S5. Change of bacterial taxa in the polyp-Y and polyp-N in biopsy samples**

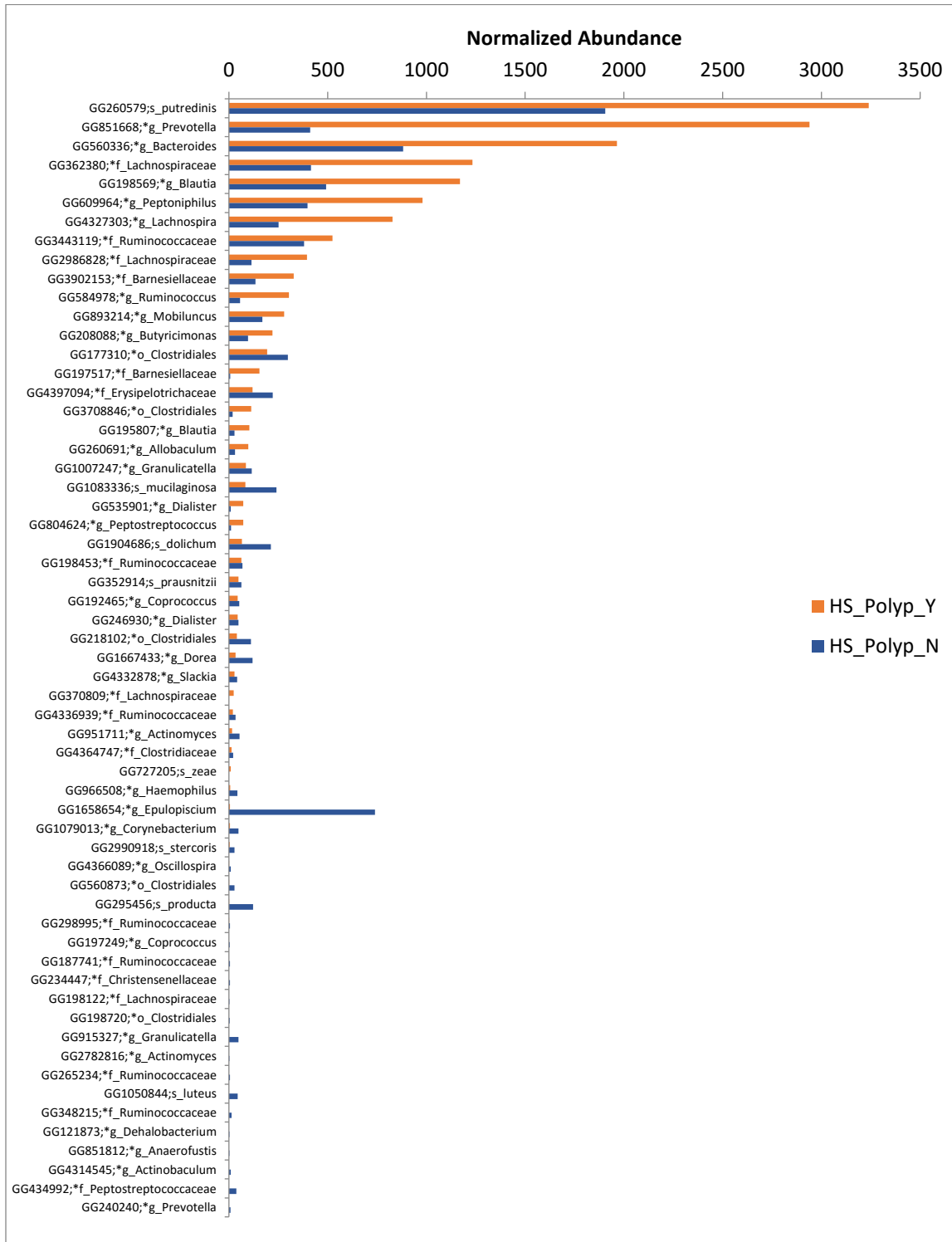




Some taxa increased, and some decreased in the polyp-Y group compared to the polyp-N.

BS, biopsy samples; Polyp-N, polyp-negative group; Polyp-Y, polyp-positive group.

**Figure S6. Change of bacterial taxa in the polyp-Y and polyp-N in home stool samples**

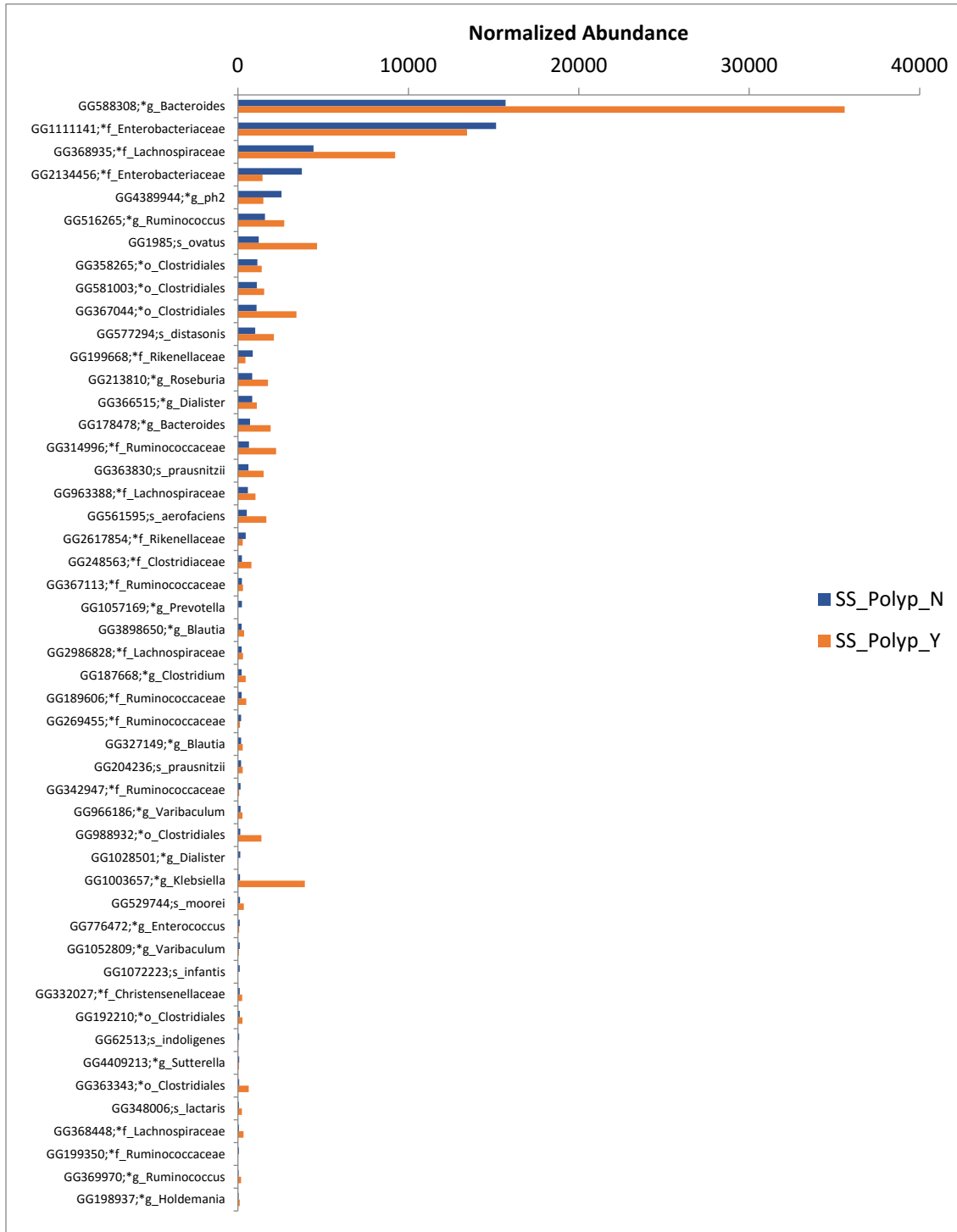


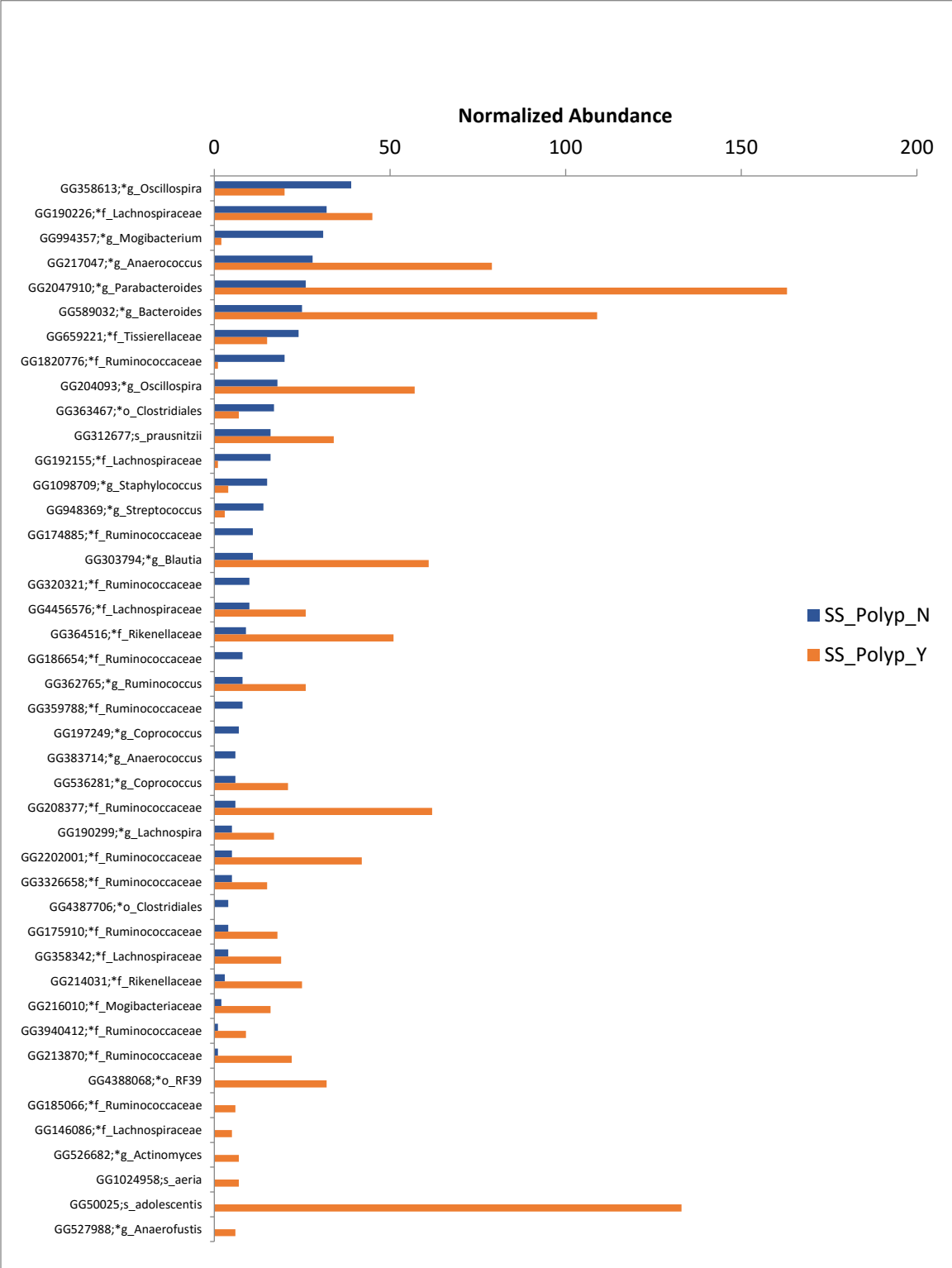
Some taxa increased, and some decreased in the polyp-Y group compared to the polyp-N.

HS, home stool samples; Polyp-N, polyp-negative group; Polyp-Y, polyp-positive group.



**Figure S7. Change of bacterial taxa in the polyp-Y and polyp-N in rectal swab samples**





Some taxa increased, and some decreased in the polyp-Y group compared to the polyp-N.

Polyp-N, polyp-negative group; Polyp-Y, polyp-positive group; SS, rectal swab samples.

## SUPPLEMENTAL TABLES

**Table S1. A summary of selected CRC and adenoma microbiome studies from 2010 to 2017**

<b>Study</b>	<b>Type of samples</b>	<b>disease</b>	<b>Population</b>	<b>Sample size</b>	<b>Method</b>	<b>Variable region</b>
<b>Shen et al. 2010</b>	Mucosal biopsy	Adenoma	USA	21 adenoma, 23 control	Terminal restriction fragment length polymorphism, clone sequencing and fluorescent in-situ hybridization analysis of the 16S rRNA genes	N/A
<b>Marchesi et al. 2011</b>	Tumor/ adjacent normal tissue	CRC	Netherlands	6 CRC	Roche 454 GS FLX pyrosequencing	V1-V3
<b>Castellarin et al. 2012</b>	Tumor/ adjacent normal tissue	CRC	Canada	11 CRC	Illumina GAIIx, RNA seq	N/A
<b>Chen et al. 2012</b>	Intestinal lumen, mucosa (rectal swabs), fecal samples, tumor/ matching normal tissue	CRC	China	46 CRC, 56 control	Roche 454 GS FLX pyrosequencing	V1-V3
<b>Kostic et al. 2012</b>	Tumor/ adjacent normal tissue	CRC	USA & Vietnam	95 CRC	Roche 454 GS FLX pyrosequencing	V3-V5
<b>Sanapareddy et al. 2012</b>	Rectal mucosa biopsy	Adenoma	USA	33 adenoma, 38 control	Roche 454 GS FLX pyrosequencing	V1-V2
<b>Geng et al. 2013</b>	Tumor/ adjacent normal tissue	CRC	China	8 CRC	Roche 454 GS FLX pyrosequencing	V1-V2
<b>McCoy et al. 2013</b>	Rectal mucosa biopsy	Adenoma	USA	48 CRC, 67 control	Roche 454 GS FLX pyrosequencing	V1-V3
<b>Zeller et al. 2014</b>	Tumor/ adjacent normal tissue	CRC	Germany	38 CRC	Illumina MiSeq	V4
<b>Burns et al. 2015</b>	Tumor/ adjacent normal tissue	CRC	USA	44 CRC	Illumina MiSeq	V5-V6
<b>Mira-Pascual et al. 2015</b>	Fecal and biopsy	CRC & adenoma	Spain	7 CRC, 11 tubular adenoma, 10 control	Roche 454 GS FLX pyrosequencing	N/A
<b>Nakatsu et al. 2015</b>	Biopsy	CRC & adenoma	China	52 Tumor/adjacent, 47 adenoma/ adjacent, 61 control	Roche 454 GS FLX pyrosequencing	V1-V4
<b>Thomas et al. 2016</b>	Biopsy	CRC	Brazil	18 rectal cancer, 18 control	Ion-torrent PGM platform	V4-V5
<b>Xu &amp; Jiang 2017</b>	Biopsy	CRC & adenoma	China	52 cancer, 47 adenoma, 61 control	Roche 454 GS FLX pyrosequencing	V1-V4
<b>Gao et al. 2017</b>	Tumor/ adjacent normal tissue	CRC	China	65 CRC	Illumina MiSeq	V4

<b>Yoon et al. 2017</b>	Biopsy	CRC & adenoma	Korea	6 CRC, 6 conventional adenoma, 6 sessile serrated adenoma, 6 control	Roche 454 GS FLX pyrosequencing	V1-V3
<b>Hale et al. 2017</b>	Fecal	Adenoma	USA	233 adenoma, 547 control	Illumina MiSeq	N/A

There are many structural differences among these studies including sample type (tissue, stool, and rectal swab), population (various genetic background, different geography), and technical differences (Sequencing method, 16S rRNA primer); these differences can affect the findings of these reports and they may not be readily comparable.

CRC, colorectal cancer; N/A, not available.

**Table S2. Alpha diversity and the change of taxa abundance at phylum level between polyp datasets**

<b>Sample</b>	<b>Alpha diversity</b>	<b>Firmicutes</b>	<b>Bacteroidetes</b>	<b>Proteobacteria</b>
<b>Biopsy</b>	polyp-Y<polyp-N	polyp-Y>polyp-N	polyp-Y>polyp-N	polyp-Y<polyp-N
<b>Home stool</b>	polyp-Y<polyp-N	polyp-Y>polyp-N	polyp-Y>polyp-N	polyp-Y>polyp-N
<b>Rectal swab</b>	No change	polyp-Y>polyp-N	polyp-Y>polyp-N	polyp-Y>polyp-N

polyp-N, polyp-negative group; polyp-Y, polyp-positive group.

**Table S3. Change of taxa abundance at genera level between polyp datasets**

<b>Sample</b>	<b>Fusobacterium</b>	<b>Bacteroides</b>	<b>Blautia</b>	<b>Bifidobacterium</b>	<b>Roseburia</b>	<b>Faecalobacterium</b>
<b>Biopsy</b>	polyp-Y< polyp-N (N/S)	polyp-Y> polyp-N	polyp-Y< polyp-N	No change	polyp-Y< polyp-N (N/S)	polyp-Y< polyp-N
<b>Home stool</b>	polyp-Y< polyp-N (N/S)	polyp-Y> polyp-N	polyp-Y> polyp-N	No change	polyp-Y> polyp-N (N/S)	polyp-Y< polyp-N
<b>Rectal swab</b>	polyp-Y< polyp-N (N/S)	polyp-Y> polyp-N	polyp-Y> polyp-N	No change	polyp-Y> polyp-N (N/S)	polyp-Y> polyp-N

N/S, not significant; polyp-N, polyp-negative group; polyp-Y, polyp-positive group

**Table S4. The informative OTUs that were significantly different between the polyp-Y and polyp-N groups in biopsy samples**

No.	OTU Number	GreenGenes ID	Taxonomy
1	OTU1005	GG349067	p_Firmicutes;c_Clostridia;o_Clostridiales;
2	OTU102	GG1067865	p_Proteobacteria;c_Betaproteobacteria;o_Burkholderiales;f_Burkholderiaceae;g_Burkholderia;
3	OTU1062	GG4315468	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;
4	OTU1063	GG1522739	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Clostridiaceae;
5	OTU1066	GG214036	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Mogibacteriaceae;
6	OTU1075	GG739971	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;g_Oscillospira;
7	OTU1130	GG213870	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;
8	OTU115	GG359930	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Blautia;s_obeum
9	OTU1177	GG517201	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;
10	OTU1202	GG4367656	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Moryella;
11	OTU1206	GG183845	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;
12	OTU1226	GG174752	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;
13	OTU1237	GG3940440	p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidaceae;g_Bacteroides;
14	OTU124	GG198054	p_Firmicutes;c_Erysipelotrichi;o_Erysipelotrichales;f_Erysipelotrichaceae;g_Eubacterium;s_dolichum
15	OTU13	GG366744	p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidaceae;g_Bacteroides;
16	OTU133	GG181572	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;
17	OTU1330	GG189667	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Dorea;
18	OTU1363	GG338730	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;
19	OTU1387	GG265234	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;
20	OTU140	GG3199564	p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Odoribacteraceae;g_Odoribacter;
21	OTU1421	GG4316515	p_Firmicutes;c_Clostridia;o_Clostridiales;
22	OTU1439	GG194172	p_Firmicutes;c_Clostridia;o_Clostridiales;
23	OTU147	GG584978	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;g_Ruminococcus;
24	OTU1490	GG608244	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;g_Ruminococcus;
25	OTU1519	GG589032	p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidaceae;g_Bacteroides;
26	OTU1520	GG187741	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;
27	OTU1533	GG520827	p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;
28	OTU154	GG332027	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Christensenellaceae;
29	OTU1546	GG296589	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;
30	OTU155	GG313166	p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidaceae;g_Bacteroides;
31	OTU1550	GG4436120	p_Firmicutes;c_Clostridia;o_Clostridiales;
32	OTU156	GG4144206	p_Firmicutes;c_Erysipelotrichi;o_Erysipelotrichales;f_Erysipelotrichaceae;
33	OTU1595	GG225338	p_Proteobacteria;c_Gammaproteobacteria;o_Pasteurellales;f_Pasteurellaceae;g_Aggregatibacter;
34	OTU1612	GG297160	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;g_Ruminococcus;
35	OTU1644	GG216862	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;
36	OTU1649	GG3804871	p_Proteobacteria;c_Gammaproteobacteria;o_Enterobacteriales;f_Enterobacteriaceae;
37	OTU1662	GG210793	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;
38	OTU1681	GG849346	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;

39	OTU1700	GG1098709	p_Firmicutes;c_Bacilli;o_Bacillales;f_Staphylococcaceae;g_Staphylococcus;
40	OTU173	GG4321260	p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Paraprevotellaceae;g_Prevotella;
41	OTU1743	GG2924870	p_Proteobacteria;c_Gammaproteobacteria;o_Enterobacteriales;f_Enterobacteriaceae;
42	OTU1747	GG725212	p_Firmicutes;c_Clostridia;o_Clostridiales;
43	OTU177	GG195088	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;g_Ruminococcus;
44	OTU1802	GG573969	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;
45	OTU1813	GG210303	p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Odoribacteraceae;g_Odoribacter;
46	OTU1886	GG320055	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;
47	OTU1950	GG331253	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Blautia;
48	OTU1975	GG519882	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;
49	OTU198	GG988932	p_Firmicutes;c_Clostridia;o_Clostridiales;
50	OTU2020	GG229348	p_Firmicutes;c_Clostridia;o_Clostridiales;
51	OTU2058	GG4311934	p_Firmicutes;c_Clostridia;o_Clostridiales;
52	OTU2185	GG560873	p_Firmicutes;c_Clostridia;o_Clostridiales;
53	OTU2194	GG1064335	p_Actinobacteria;c_Actinobacteria;o_Actinomycetales;f_Actinomycetaceae;g_Actinomyces;
54	OTU241	GG256015	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Clostridiaceae;
55	OTU252	GG198118	p_Firmicutes;c_Clostridia;o_Clostridiales;
56	OTU2522	GG194097	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;
57	OTU257	GG366987	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;g_Faecalibacterium;s_prausnitzii
58	OTU271	GG4387453	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Clostridiaceae;
59	OTU284	GG184217	p_Firmicutes;c_Clostridia;o_Clostridiales;
60	OTU318	GG2617854	p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Rikenellaceae;
61	OTU320	GG1097287	p_Firmicutes;c_Bacilli;o_Bacillales;f_Staphylococcaceae;g_Staphylococcus;s_epidermidis
62	OTU334	GG181826	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;
63	OTU366	GG516265	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Ruminococcus;
64	OTU384	GG16274	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Clostridiaceae;g_Clostridium;
65	OTU394	GG198085	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;g_Faecalibacterium;s_prausnitzii
66	OTU398	GG192155	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;
67	OTU404	GG308631	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;
68	OTU415	GG324015	p_Firmicutes;c_Clostridia;o_Clostridiales;
69	OTU436	GG1024958	p_Actinobacteria;c_Actinobacteria;o_Actinomycetales;f_Micrococcaceae;g_Rothia;s_aeria
70	OTU444	GG204236	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;g_Faecalibacterium;s_prausnitzii
71	OTU449	GG198937	p_Firmicutes;c_Erysipelotrichi;o_Erysipelotrichales;f_Erysipelotrichaceae;g_Holdemania;
72	OTU461	GG359359	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;
73	OTU474	GG330439	p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Prevotellaceae;g_Prevotella;s_stercora
74	OTU475	GG354574	p_Proteobacteria;c_Deltaproteobacteria;o_Desulfovibrionales;f_Desulfovibrionaceae;g_Bilophila;
75	OTU51	GG1003657	p_Proteobacteria;c_Gammaproteobacteria;o_Enterobacteriales;f_Enterobacteriaceae;g_Klebsiella;
76	OTU514	GG848116	p_Proteobacteria;c_Betaproteobacteria;o_Burkholderiales;f_Alcaligenaceae;g_Sutterella;
77	OTU550	GG2134456	p_Proteobacteria;c_Gammaproteobacteria;o_Enterobacteriales;f_Enterobacteriaceae;
78	OTU553	GG1135793	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Peptococcaceae;g_Peptococcus;
79	OTU555	GG199350	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;



80	OTU567	GG585220	p_Firmicutes;c_Erysipelotrichi;o_Erysipelotrichales;f_Erysipelotrichaceae;g_cc_115;
81	OTU571	GG187802	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Mogibacteriaceae;
82	OTU576	GG218102	p_Firmicutes;c_Clostridia;o_Clostridiales;
83	OTU6	GG1111141	p_Proteobacteria;c_Gammaproteobacteria;o_Enterobacteriales;f_Enterobacteriaceae;
84	OTU600	GG2042960	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Tissierellaceae;g_Peptoniphilus;
85	OTU612	GG531950	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;
86	OTU647	GG1007247	p_Firmicutes;c_Bacilli;o_Lactobacillales;f_Carnobacteriaceae;g_Granulicatella;
87	OTU649	GG353631	p_Firmicutes;c_Clostridia;o_Clostridiales;
88	OTU65	GG305288	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Clostridiaceae;g_Clostridium;
89	OTU655	GG4315787	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;
90	OTU659	GG176104	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;g_Oscillospira;
91	OTU662	GG179815	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Ruminococcus;s_gnavus
92	OTU697	GG4303851	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;g_Ruminococcus;
93	OTU723	GG230405	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Christensenellaceae;
94	OTU737	GG4470076	p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Prevotellaceae;g_Prevotella;s_nigrescens
95	OTU76	GG851704	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Tissierellaceae;g_Parvimonas;
96	OTU761	GG358030	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;
97	OTU791	GG4453550	p_Firmicutes;c_Erysipelotrichi;o_Erysipelotrichales;f_Erysipelotrichaceae;g_RFN20;
98	OTU812	GG191100	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;
99	OTU82	GG192424	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Blautia;
100	OTU83	GG882886	p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidaceae;g_Bacteroides;s_fragilis
101	OTU838	GG20310	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Mogibacteriaceae;
102	OTU844	GG4440820	p_Fusobacteria;c_Fusobacteriia;o_Fusobacteriales;f_Leptotrichiaceae;g_Leptotrichia;
103	OTU85	GG843886	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;
104	OTU861	GG4329995	p_Proteobacteria;c_Deltaproteobacteria;o_Desulfovibrionales;f_Desulfovibrionaceae;
105	OTU874	GG932843	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Tissierellaceae;g_Anaerococcus;
106	OTU899	GG105514	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;
107	OTU905	GG575852	p_Proteobacteria;c_Gammaproteobacteria;o_Enterobacteriales;f_Enterobacteriaceae;g_Citrobacter;
108	OTU914	GG1835985	p_Firmicutes;c_Clostridia;o_Clostridiales;
109	OTU969	GG574988	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;

From 109 OTUs, 83 of them belong to Firmicutes phylum, and 12 OTUs were classified as Bacteroidetes Phylum; 10 OTUs have taxonomies up to species level.

c, class; f, family; g, genus; o, order; OTU, operational taxonomic unit; p, phylum; polyp-N, polyp-negative group; polyp-Y, polyp-positive group; s, species.

**Table S5. The informative OTUs that were significantly different between the polyp-Y and polyp-N groups in the home stool samples**

No.	OTU Number	GreenGenes ID	Taxonomy
1	OTU1002	GG915327	p_Firmicutes;c_Bacilli;o_Lactobacillales;f_Carnobacteriaceae;g_Granulicatella;
2	OTU1159	GG295456	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Blautia;s_producta
3	OTU1166	GG727205	p_Firmicutes;c_Bacilli;o_Lactobacillales;f_Lactobacillaceae;g_Lactobacillus;s_zeae
4	OTU12	GG260579	p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Rikenellaceae;g_Alistipes;s_putredinis
5	OTU125	GG3443119	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;
6	OTU126	GG3902153	p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Barnesiellaceae;
7	OTU1353	GG1904686	p_Firmicutes;c_Erysipelotrichi;o_Erysipelotrichales;f_Erysipelotrichaceae;g_Eubacterium;s_dolichum
8	OTU1374	GG4332878	p_Actinobacteria;c_Coriobacteriia;o_Coriobacteriales;f_Coriobacteriaceae;g_Slackia;
9	OTU1387	GG265234	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;
10	OTU147	GG584978	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;g_Ruminococcus;
11	OTU1476	GG851812	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Eubacteriaceae;g_Anaerofustis;
12	OTU1520	GG187741	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;
13	OTU1576	GG121873	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Dehalobacteriaceae;g_Dehalobacterium;
14	OTU1608	GG234447	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Christensenellaceae;
15	OTU1774	GG260691	p_Firmicutes;c_Erysipelotrichi;o_Erysipelotrichales;f_Erysipelotrichaceae;g_Allobaculum;
16	OTU1899	GG348215	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;
17	OTU1942	GG197249	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Coproccoccus;
18	OTU2022	GG2782816	p_Actinobacteria;c_Actinobacteria;o_Actinomycetales;f_Actinomycetaceae;g_Actinomyces;
19	OTU2159	GG4366089	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;g_Oscillospira;
20	OTU2221	GG240240	p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Paraprevotellaceae;g_Prevotella;
21	OTU2427	GG198720	p_Firmicutes;c_Clostridia;o_Clostridiales;
22	OTU243	GG362380	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;
23	OTU2495	GG198122	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;
24	OTU2525	GG2990918	p_Actinobacteria;c_Coriobacteriia;o_Coriobacteriales;f_Coriobacteriaceae;g_Collinsella;s_stercoris
25	OTU258	GG198569	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Blautia;
26	OTU2637	GG1079013	p_Actinobacteria;c_Actinobacteria;o_Actinomycetales;f_Corynebacteriaceae;g_Corynebacterium;
27	OTU274	GG1083336	p_Actinobacteria;c_Actinobacteria;o_Actinomycetales;f_Micrococcaceae;g_Rothia;s_mucilaginoso
28	OTU305	GG197517	p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Barnesiellaceae;
29	OTU319	GG4397094	p_Firmicutes;c_Erysipelotrichi;o_Erysipelotrichales;f_Erysipelotrichaceae;
30	OTU324	GG4327303	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Lachnospira;
31	OTU327	GG3708846	p_Firmicutes;c_Clostridia;o_Clostridiales;
32	OTU346	GG1658654	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Epulopiscium;
33	OTU35	GG851668	p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Prevotellaceae;g_Prevotella;
34	OTU352	GG246930	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Veillonellaceae;g_Dialister;
35	OTU389	GG208088	p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Odoribacteraceae;g_Butyricimonas;
36	OTU409	GG1667433	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Dorea;
37	OTU434	GG966508	p_Proteobacteria;c_Gammaproteobacteria;o_Pasteurellales;f_Pasteurellaceae;g_Haemophilus;
38	OTU440	GG177310	p_Firmicutes;c_Clostridia;o_Clostridiales;

39	OTU443	GG198453	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;
40	OTU453	GG2986828	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;
41	OTU517	GG535901	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Veillonellaceae;g_Dialister;
42	OTU535	GG195807	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Blautia;
43	OTU560	GG804624	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Peptostreptococcaceae;g_Peptostreptococcus;
44	OTU576	GG218102	p_Firmicutes;c_Clostridia;o_Clostridiales;
45	OTU593	GG434992	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Peptostreptococcaceae;
46	OTU636	GG4336939	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;
47	OTU647	GG1007247	p_Firmicutes;c_Bacilli;o_Lactobacillales;f_Carnobacteriaceae;g_Granulicatella;
48	OTU675	GG893214	p_Actinobacteria;c_Actinobacteria;o_Actinomycetales;f_Actinomycetaceae;g_Mobiluncus;
49	OTU733	GG1050844	p_Actinobacteria;c_Actinobacteria;o_Actinomycetales;f_Micrococcaceae;g_Micrococcus;s_luteus
50	OTU747	GG560336	p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidaceae;g_Bacteroides;
51	OTU785	GG192465	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Coprococcus;
52	OTU786	GG951711	p_Actinobacteria;c_Actinobacteria;o_Actinomycetales;f_Actinomycetaceae;g_Actinomyces;
53	OTU790	GG560873	p_Firmicutes;c_Clostridia;o_Clostridiales;
54	OTU808	GG4314545	p_Actinobacteria;c_Actinobacteria;o_Actinomycetales;f_Actinomycetaceae;g_Actinobaculum;
55	OTU819	GG4364747	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Clostridiaceae;
56	OTU847	GG370809	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;
57	OTU96	GG609964	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Tissierellaceae;g_Peptoniphilus;
58	OTU960	GG298995	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;
59	OTU964	GG352914	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;g_Faecalibacterium;s_prausnitzii

From 59 informative OTUs, 42 of them belong to Firmicutes phylum, 7 OTUs to Bacteroidetes phylum; the taxonomy of 5 OTUs is clear up to species level.

c, class; f, family; g, genus; o, order; OTU, operational taxonomic unit; p, phylum; polyp-N, polyp-negative group; polyp-Y, polyp-positive group; s, species.

**Table S6. The informative OTUs that were significantly different between the polyp-Y and polyp-N groups in the rectal swab samples**

No.	OTU Number	GreenGenes ID	Taxonomy
1	OTU1001	GG4456576	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;
2	OTU1014	GG178478	p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidaceae;g_Bacteroides;
3	OTU1033	GG189606	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;
4	OTU106	GG366515	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Veillonellaceae;g_Dialister;
5	OTU1071	GG174885	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;
6	OTU1098	GG4388068	p_Tenericutes;c_Mollicutes;o_RF39;
7	OTU112	GG187668	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Clostridiaceae;g_Clostridium;
8	OTU116	GG314996	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;
9	OTU1164	GG994357	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Mogibacteriaceae;g_Mogibacterium;
10	OTU1167	GG359788	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;
11	OTU117	GG963388	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;
12	OTU122	GG248563	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Clostridiaceae;
13	OTU1298	GG213870	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;
14	OTU1453	GG146086	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;
15	OTU149	GG1057169	p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Prevotellaceae;g_Prevotella;
16	OTU1513	GG3940412	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;
17	OTU154	GG332027	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Christensenellaceae;
18	OTU1558	GG269455	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;
19	OTU158	GG776472	p_Firmicutes;c_Bacilli;o_Lactobacillales;f_Enterococcaceae;g_Enterococcus;
20	OTU1584	GG185066	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;
21	OTU1611	GG3326658	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;
22	OTU1633	GG948369	p_Firmicutes;c_Bacilli;o_Lactobacillales;f_Streptococcaceae;g_Streptococcus;
23	OTU1700	GG1098709	p_Firmicutes;c_Bacilli;o_Bacillales;f_Staphylococcaceae;g_Staphylococcus;
24	OTU1707	GG208377	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;
25	OTU1732	GG4387706	p_Firmicutes;c_Clostridia;o_Clostridiales;
26	OTU175	GG4409213	p-Proteobacteria;c_Betaproteobacteria;o_Burkholderiales;f_Alcaligenaceae;g_Sutterella;
27	OTU1781	GG358342	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;
28	OTU193	GG367113	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;
29	OTU1942	GG197249	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Coproccoccus;
30	OTU198	GG988932	p_Firmicutes;c_Clostridia;o_Clostridiales;
31	OTU1992	GG190299	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Lachnospira;
32	OTU2	GG588308	p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidaceae;g_Bacteroides;
33	OTU2001	GG526682	p_Actinobacteria;c_Actinobacteria;o_Actinomycetales;f_Actinomycetaceae;g_Actinomyces;
34	OTU2068	GG527988	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Eubacteriaceae;g_Anaerofustis;
35	OTU226	GG358265	p_Firmicutes;c_Clostridia;o_Clostridiales;
36	OTU2283	GG348006	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Ruminococcus;s_lactaris
37	OTU2378	GG364516	p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Rikenellaceae;
38	OTU2413	GG3898650	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Blautia;

39	OTU2447	GG312677	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;g_Faecalibacterium;s_prausnitzii
40	OTU2536	GG1072223	p_Firmicutes;c_Bacilli;o_Lactobacillales;f_Streptococcaceae;g_Streptococcus;s_infantis
41	OTU2565	GG589032	p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidaceae;g_Bacteroides;
42	OTU27	GG4389944	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Tissierellaceae;g_ph2;
43	OTU272	GG363830	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;g_Faecalibacterium;s_prausnitzii
44	OTU286	GG199668	p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Rikenellaceae;
45	OTU304	GG369970	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Ruminococcus;
46	OTU31	GG213810	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Roseburia;
47	OTU318	GG2617854	p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Rikenellaceae;
48	OTU325	GG1028501	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Veillonellaceae;g_Dialister;
49	OTU328	GG50025	p_Actinobacteria;c_Actinobacteria;o_Bifidobacteriales;f_Bifidobacteriaceae;g_Bifidobacterium;s_adolescentis
50	OTU329	GG214031	p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Rikenellaceae;
51	OTU331	GG368448	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;
52	OTU336	GG62513	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Moryella;s_indoligenes
53	OTU362	GG966186	p_Actinobacteria;c_Actinobacteria;o_Actinomycetales;f_Actinomycetaceae;g_Varibaculum;
54	OTU366	GG516265	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Ruminococcus;
55	OTU368	GG204093	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;g_Oscillospira;
56	OTU398	GG192155	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;
57	OTU436	GG1024958	p_Actinobacteria;c_Actinobacteria;o_Actinomycetales;f_Micrococcaceae;g_Rothia;s_aeria
58	OTU444	GG204236	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;g_Faecalibacterium;s_prausnitzii
59	OTU449	GG198937	p_Firmicutes;c_Erysipelotrichi;o_Erysipelotrichales;f_Erysipelotrichaceae;g_Holdemania;
60	OTU45	GG581003	p_Firmicutes;c_Clostridia;o_Clostridiales;
61	OTU453	GG2986828	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;
62	OTU454	GG2202001	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;
63	OTU457	GG358613	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;g_Oscillospira;
64	OTU469	GG383714	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Tissierellaceae;g_Anaerococcus;
65	OTU470	GG1052809	p_Actinobacteria;c_Actinobacteria;o_Actinomycetales;f_Actinomycetaceae;g_Varibaculum;
66	OTU473	GG192210	p_Firmicutes;c_Clostridia;o_Clostridiales;
67	OTU496	GG342947	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;
68	OTU5	GG368935	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;
69	OTU50	GG529744	p_Firmicutes;c_Erysipelotrichi;o_Erysipelotrichales;f_Erysipelotrichaceae;g_Bulleidia;s_moorei
70	OTU51	GG1003657	p-Proteobacteria;c_Gammaproteobacteria;o_Enterobacteriales;f_Enterobacteriaceae;g_Klebsiella;
71	OTU520	GG659221	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Tissierellaceae;
72	OTU53	GG367044	p_Firmicutes;c_Clostridia;o_Clostridiales;
73	OTU550	GG2134456	p-Proteobacteria;c_Gammaproteobacteria;o_Enterobacteriales;f_Enterobacteriaceae;
74	OTU555	GG199350	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;
75	OTU59	GG577294	p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Porphyrimonadaceae;g_Parabacteroides;s_distasonis
76	OTU6	GG1111141	p-Proteobacteria;c_Gammaproteobacteria;o_Enterobacteriales;f_Enterobacteriaceae;
77	OTU628	GG2047910	p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Porphyrimonadaceae;g_Parabacteroides;
78	OTU63	GG363343	p_Firmicutes;c_Clostridia;o_Clostridiales;
79	OTU632	GG217047	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Tissierellaceae;g_Anaerococcus;

80	OTU672	GG362765	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;g_Ruminococcus;
81	OTU686	GG327149	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Blautia;
82	OTU720	GG175910	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;
83	OTU739	GG303794	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Blautia;
84	OTU768	GG1820776	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;
85	OTU77	GG1985	p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidaceae;g_Bacteroides;s_ovatus
86	OTU772	GG363467	p_Firmicutes;c_Clostridia;o_Clostridiales;
87	OTU846	GG186654	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;
88	OTU849	GG190226	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;
89	OTU872	GG561595	p_Actinobacteria;c_Coriobacteriia;o_Coriobacteriales;f_Coriobacteriaceae;g_Collinsella;s_aerofaciens
90	OTU889	GG320321	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;
91	OTU929	GG216010	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Mogibacteriaceae;
92	OTU966	GG536281	p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Coprococcus;

From 92 informative OTUs, 69 OTUs belong to Firmicutes, and 12 OTUs belong to Bacteroidetes Phyla; the taxonomy of 10 OTUs is clear up to species level.

c, class; f, family; g, genus; o, order; OTU, operational taxonomic unit; p, phylum; polyp-N, polyp-negative group; polyp-Y, polyp-positive group; s, species.

**Table S7. The classification validation results of polyp dataset**

Raw_BS					Raw_HS					Raw_SS				
Classification method	CA	Sens	Spec	AUC	Classification method	CA	Sens	Spec	AUC	Classification method	CA	Sens	Spec	AUC
Naïve Bayes	0.43	0	1	0.59	Naïve Bayes	0.40	0	1	0.57	Naïve Bayes	0.44	0	0.98	0.57
Random Forest	0.63	0.77	0.44	0.66	Random Forest	0.55	0.78	0.23	0.56	Random Forest	0.59	0.69	0.47	0.65
kNN	0.55	0.66	0.39	0.61	kNN	0.58	0.87	0.16	0.52	kNN	0.49	0.73	0.21	0.57
Classification Tree	0.56	0.66	0.41	0.57	Classification Tree	0.59	0.67	0.49	0.55	Classification Tree	0.54	0.55	0.53	0.55
Logistic regression	0.56	0.59	0.51	0.57	Logistic regression	0.52	0.59	0.41	0.52	Logistic regression	0.56	0.55	0.57	0.56
Neural Network	0.56	0.63	0.46	0.57	Neural Network	0.50	0.60	0.36	0.53	Neural Network	0.55	0.57	0.52	0.56
SVM	0.57	1	0	0.50	SVM	0.59	1	0	0.5	SVM	0.54	1	0	0.49
Filtered_BS					Filtered_HS					Filtered_SS				
Classification method	CA	Sens	Spec	AUC	Classification method	CA	Sens	Spec	AUC	Classification method	CA	Sens	Spec	AUC
Naïve Bayes	0.77	0.78	0.74	0.85	Naïve Bayes	0.75	0.73	0.78	0.83	Naïve Bayes	0.72	0.71	0.73	0.80
Random Forest	0.73	0.80	0.62	0.79	Random Forest	0.72	0.88	0.49	0.83	Random Forest	0.69	0.76	0.60	0.81
kNN	0.72	0.87	0.51	0.76	kNN	0.66	0.81	0.45	0.70	kNN	0.61	0.63	0.60	0.66
Classification Tree	0.56	0.63	0.46	0.50	Classification Tree	0.64	0.69	0.56	0.62	Classification Tree	0.57	0.61	0.51	0.57
Logistic regression	0.76	0.80	0.69	0.81	Logistic regression	0.79	0.86	0.69	0.81	Logistic regression	0.70	0.71	0.68	0.77
Neural Network	0.74	0.80	0.65	0.81	Neural Network	0.77	0.84	0.67	0.81	Neural Network	0.67	0.69	0.65	0.76
SVM	0.6	0.98	0.09	0.69	SVM	0.70	0.73	0.65	0.74	SVM	0.60	0.88	0.26	0.68
On test data					On test data					On test data				
Classification method	CA	Sens	Spec	AUC	Classification method	CA	Sens	Spec	AUC	Classification method	CA	Sens	Spec	AUC
Naïve Bayes	0.88	1	0.72	0.91	Naïve Bayes	0.85	0.85	0.85	0.94	Naïve Bayes	0.69	0.69	0.7	0.72
Random Forest	0.6	0.71	0.45	0.72	Random Forest	0.76	0.9	0.57	0.90	Random Forest	0.67	0.69	0.65	0.71
kNN	0.68	0.85	0.45	0.75	kNN	0.67	0.95	0.28	0.73	kNN	0.58	0.56	0.6	0.63
Classification Tree	0.48	0.5	0.45	0.44	Classification Tree	0.67	0.7	0.64	0.70	Classification Tree	0.62	0.56	0.7	0.61
Logistic regression	0.88	1	0.72	0.95	Logistic regression	0.91	0.9	0.92	0.94	Logistic regression	0.60	0.65	0.55	0.65
Neural Network	0.84	1	0.63	0.95	Neural Network	0.91	0.9	0.92	0.94	Neural Network	0.58	0.65	0.5	0.67
SVM	0.52	0.92	0	0.69	SVM	0.61	0.75	0.42	0.76	SVM	0.58	0.86	0.25	0.58

The top tables show validation results of the classifiers produced using all OTUs. The middle tables used just the informative OTUs as classification features. The bottom tables are the result of applying the second classifier on the test dataset. Classifiers perform better when informative OTUs are used as classification features.

AUC, area under the curve; BS, biopsy sample; CA, classification accuracy; HS, home stool sample; KNN, K-nearest neighbor; OTU, Operational Taxonomic Unit; SS, rectal swab sample; Sens, sensitivity; Spec, specificity; SVM, support vector machine.

**Table S8. Polyp prediction using the biopsy dataset**

<b>Sample</b>	<b>Actual Class</b>	<b>Predicted class using Naïve Bayes classifier</b>
<b>BS_9</b>	BS_Polyp_Y	BS_Polyp_Y
<b>BS_419</b>	BS_Polyp_Y	BS_Polyp_Y
<b>BS_353</b>	BS_Polyp_Y	BS_Polyp_N
<b>BS_431</b>	BS_Polyp_Y	BS_Polyp_N
<b>BS_436</b>	BS_Polyp_Y	BS_Polyp_Y
<b>BS_405</b>	BS_Polyp_Y	BS_Polyp_Y
<b>BS_351</b>	BS_Polyp_Y	BS_Polyp_Y
<b>BS_389</b>	BS_Polyp_Y	BS_Polyp_Y
<b>BS_57</b>	BS_Polyp_N	BS_Polyp_Y
<b>BS_517</b>	BS_Polyp_N	BS_Polyp_Y
<b>BS_330</b>	BS_Polyp_N	BS_Polyp_N
<b>BS_344</b>	BS_Polyp_N	BS_Polyp_Y
<b>BS_421</b>	BS_Polyp_N	BS_Polyp_N

Correctly classified biopsy test samples are 61.5% of the samples.

BS, biopsy sample; Polyp-N, polyp-negative group; Polyp-Y, polyp-positive group.



**Table S9. Polyp prediction using the home stool dataset**

<b>Sample</b>	<b>Actual Class</b>	<b>Predicted class using Naïve Bayes classifier</b>
<b>HS_458</b>	HS_Polyp_N	HS_Polyp_N
<b>HS_20</b>	HS_Polyp_N	HS_Polyp_N
<b>HS_396</b>	HS_Polyp_N	HS_Polyp_N
<b>HS_370</b>	HS_Polyp_N	HS_Polyp_N
<b>HS_399</b>	HS_Polyp_N	HS_Polyp_N
<b>HS_511</b>	HS_Polyp_N	HS_Polyp_N
<b>HS_2</b>	HS_Polyp_N	HS_Polyp_Y
<b>HS_314</b>	HS_Polyp_Y	HS_Polyp_Y
<b>HS_62</b>	HS_Polyp_Y	HS_Polyp_Y
<b>HS_391</b>	HS_Polyp_Y	HS_Polyp_Y
<b>HS_303</b>	HS_Polyp_Y	HS_Polyp_Y
<b>HS_311</b>	HS_Polyp_Y	HS_Polyp_Y
<b>HS_318</b>	HS_Polyp_Y	HS_Polyp_Y
<b>HS_518</b>	HS_Polyp_Y	HS_Polyp_Y
<b>HS_449</b>	HS_Polyp_Y	HS_Polyp_Y
<b>HS_6</b>	HS_Polyp_Y	HS_Polyp_N
<b>HS_23</b>	HS_Polyp_Y	HS_Polyp_N

Correctly classified stool test samples are 82.3% of the samples.

HS, home stool sample; Polyp-N, polyp-negative group; Polyp-Y, polyp-positive group.

**Table S10. Polyp prediction using the rectal swab dataset**

<b>Sample</b>	<b>Actual Class</b>	<b>Predicted class using Naïve Bayes classifier</b>
SS_3	SS_Polyp_N	SS_Polyp_N
SS_31	SS_Polyp_N	SS_Polyp_N
SS_469	SS_Polyp_N	SS_Polyp_N
SS_429	SS_Polyp_N	SS_Polyp_N
SS_24	SS_Polyp_N	SS_Polyp_N
SS_5	SS_Polyp_N	SS_Polyp_N
SS_302	SS_Polyp_N	SS_Polyp_Y
SS_25	SS_Polyp_N	SS_Polyp_N
SS_457	SS_Polyp_N	SS_Polyp_N
SS_57	SS_Polyp_N	SS_Polyp_N
SS_51	SS_Polyp_Y	SS_Polyp_Y
SS_347	SS_Polyp_Y	SS_Polyp_N
SS_516	SS_Polyp_Y	SS_Polyp_Y
SS_352	SS_Polyp_Y	SS_Polyp_Y
SS_34	SS_Polyp_Y	SS_Polyp_Y
SS_420	SS_Polyp_Y	SS_Polyp_Y
SS_362	SS_Polyp_Y	SS_Polyp_N
SS_16	SS_Polyp_Y	SS_Polyp_Y
SS_328	SS_Polyp_Y	SS_Polyp_N
SS_8	SS_Polyp_Y	SS_Polyp_Y
SS_53	SS_Polyp_Y	SS_Polyp_Y
SS_356	SS_Polyp_Y	SS_Polyp_Y

Correctly classified rectal swab test samples are 81.8% of the samples.

Polyp-N, polyp-negative group; Polyp-Y, polyp-positive group; SS, rectal swab sample.

**Table S11. Comparison of the alpha diversity and the change of taxa abundance at phylum level in different adenoma and CRC studies**

Study	Samples	Disease state	Alpha diversity	Firmicutes	Bacteroidetes	Proteobacteria
Shen et al. 2010	Mucosal biopsy	Adenoma	Ad>H	N/S	Ad<H	Ad>H
Marchesi et al. 2011	Tumor/adjacent normal tissue	CRC	N/A	CRC<H	CRC>H	N/A
Chen et al. 2012	rectal swabs, fecal samples, tumor/matching normal tissue	CRC	Tissue: Tumor< normal	No change	Tumor> normal	Tumor< normal
			Swab: CRC<H	No change	CRC<H	
			Stool: CRC<H	No change	CRC>H	
Kostic et al. 2012	Tumor/matching normal	CRC	N/A	Tumor< normal	Tumor< normal	N/A
Sanapareddy et al. 2012	Rectal mucosa biopsy	Adenoma	Ad>H	Ad>H	Ad>H	Ad>H
Wu et al. 2013	Fecal	CRC	N/S	CRC<H	CRC>H	N/S
McCoy et al. 2013	Rectal mucosa biopsy	Adenoma	CRC>H	N/A	CRC<H	N/A
Brim et al. 2013	Fecal	Adenoma	N/A	Ad>H	Ad<H	Ad>H
Ahn et al. 2013	Fecal	CRC	CRC<H	CRC<H	CRC>H	N/A
Zackular et al. 2014	Fecal	CRC & adenoma	N/A	N/A	N/A	N/A
Mira-Pascual et al. 2015	Fecal and biopsy	CRC & adenoma	CRC>Ad>H	N/A	Biopsy: CRC & Ad>H	N/A
Goedert et al. 2015	Fecal	CRC & adenoma	No change	Ad<H	Ad>H	Ad>H
Thomas et al., 2016	Biopsy	CRC	CRC>H	N/A	CRC>H	CRC<H
Xu and Jiang 2017	Biopsy	CRC & adenoma	Ad<H (N/S) CRC>H (N/S)	Ad<H (N/S) CRC>H	N/A	CRC<H Ad>H
Gao et al. 2017	Tumor/matching normal	CRC	No significant change	Tumor< normal	Tumor> normal	Tumor> normal
Yoon et al. 2017	Biopsy	CRC & adenoma	CRC<H Ad<H	CRC<H Ad>H	CRC<H Ad>H	CRC>H Ad<H
Hale et al. 2017	Fecal	Adenoma	No change	Ad<H	Ad>H	N/A
Our polyp dataset 2016	Fecal/ rectal swab/ biopsy	Polyp	Biopsy: Polyp-Y<polyp-N	polyp-Y>Polyp-N	polyp-Y>Polyp-N	polyp-Y<Polyp-N
			Swab: No change Polyp-Y<polyp-N	polyp-Y>Polyp-N	polyp-Y>Polyp-N	polyp-Y>Polyp-N
			stool: Polyp-Y<polyp-N	polyp-Y>Polyp-N	polyp-Y>Polyp-N	polyp-Y>Polyp-N

There is no consensus among various studies in the alpha diversity and bacterial abundance at phylum level for Bacteroidetes, Firmicutes, and Proteobacteria. In most of the studies, alpha diversity increased in the adenoma and cancer group. Proteobacteria has been reported to increase in the adenoma/CRC state in most of the studies. Ad, Adenoma; CRC, colorectal cancer; H, Healthy; N/A, not available; N/S, not significant.

**Table S12. Comparison of the frequently reported genera between our polyp dataset and seventeen previous adenoma/CRC studies**

Study	Samples	Disease state	Fusobacterium	Bacteroides	Blautia	Bifidobacterium	Roseburia	Faecalobacterium
Shen et al. 2010	Mucosal biopsy	Adenoma	N/A	Ad<H	N/A	N/A	N/A	Ad>H
Marchesi et al. 2011	Tumor/adjacent normal tissue	CRC	Tumor> normal	N/A	N/A	N/A	Tumor> normal	CRC>H
Chen et al. 2012	tumor/ matching normal tissue	CRC	Tissue: N/A	Tumor> normal	Tumor< normal	N/A	Tumor< normal	CRC>H
	Intestinal lumen, mucosa (rectal swabs), fecal samples,		Swab: CRC>H	N/A	CRC<H	CRC<H	N/A	CRC<H
			Stool: N/A	N/A	N/A	N/A	N/A	N/A
Kostic et al. 2012	Tumor/matching normal	CRC	Tumor> normal	N/A	N/A	N/A	N/A	N/A
Sanapareddy et al. 2012	Rectal mucosa biopsy	Adenoma	N/A	N/A	N/A	N/A	N/A	N/A
Wu et al. 2013	Fecal	CRC	CRC>H	CRC>H	CRC>H	N/A	CRC<H	CRC<H
McCoy et al. 2013	Rectal mucosa biopsy	Adenoma	Ad>H	N/A	N/A	N/A	N/A	N/A
Brim et al. 2013	Fecal	Adenoma	N/A	Ad<H	N/A	N/A	No change	Ad>H
Ahn et al. 2013	Fecal	CRC	CRC>H	N/A	N/A	N/A	N/A	N/A
Zackular et al. 2014	Fecal	CRC & adenoma	CRC>H	Ad<H CRC<H	Ad>CRC	N/A	N/A	N/A
Mira-Pascual et al. 2015	Fecal and biopsy	CRC & adenoma	Mucosal: CRC>H	Stool: CRC>H	Biopsy: Ad>H CRC>H	Stool: Ad<H CRC<H	N/A	N/A
Goedert et al. 2015	Fecal	CRC & Adenoma	Ad<H	Ad<H Ad>H	Ad<H Ad>H	N/A	N/A	N/A
Thomas et al., 2016	Biopsy	CRC	CRC>H	CRC>H	N/A	N/A	CRC>H	N/A
Xu and Jiang 2017	Biopsy	CRC & adenoma	CRC>H Ad<H	CRC>Ad>H	CRC<H	N/A	N/A	CRC<H
Gao et al. 2017	Tumor/matching normal	CRC	Tumor> normal	Tumor> normal	N/A	N/A	N/A	Tumor< normal
Yoon et al. 2017	Biopsy	CRC & adenoma	N/A	CRC<H Ad>H	CRC<H Ad>H	N/A	N/A	CRC<H Ad>H
Hale et al. 2017	Fecal	CRC	No enrichment in adenoma	N/A	N/A	N/A	N/A	N/A
Our polyp dataset 2016	Fecal/ swab/ biopsy	Polyp	Biopsy: polyp-Y<polyp-N (N/S)	polyp-Y> polyp-N	polyp-Y< polyp-N	No change	polyp-Y< polyp-N (N/S)	polyp-Y< polyp-N
			Swab: polyp-Y< polyp-N (N/S)	polyp-Y> polyp-N	polyp-Y> polyp-N	No change	polyp-Y> polyp-N (N/S)	polyp-Y> polyp-N
			Stool: polyp-Y< polyp-N (N/S)	polyp-Y> polyp-N	polyp-Y> polyp-N	No change	polyp-Y> polyp-N (N/S)	polyp-Y< polyp-N

For most of these taxa, no consensus conclusion of enrichment or depletion can be deduced from these studies. Fusobacterium showed enrichment in all CRC groups, however, in adenoma, it showed mostly depletion than enrichment in the disease state. Bifidobacterium reported to significantly decrease in the CRC and adenoma groups in two studies. However, other studies have not reported these taxa as a significant feature.

Ad, adenoma; CRC, colorectal cancer; H, healthy; N/A, not available; N/S, not significant.

## Differential taxa

***Fusobacterium***: ten of seventeen studies summarized in Table S12 reported an increase of *Fusobacterium* abundance in cancer patients compared to healthy control subjects (Marchesi et al., 2011; Chen et al., 2012; Kostic et al., 2012; Wu et al., 2013; Ahn et al., 2013; Zackular et al., 2014; Mira-Pascual et al., 2015; Thomas et al., 2016; Xu & Jiang, 2017; Gao et al., 2017). Among nine adenoma studies listed here, only three reported associations of adenomatous growth with *Fusobacterium* (McCoy et al., 2013; Goedert et al., 2015; Xu & Jiang, 2017). McCoy et al., observed that *Fusobacterium* abundance was significantly higher in adenomas as compared to non-adenoma samples. In their study, a high level of *Fusobacterium* increases the chance of finding adenoma by 3.5 fold. Moreover, a positive correlation between *Fusobacterium* presence and cytokine levels was reported in the adenoma cohort. Specifically, a significant association between the levels of TNF- $\alpha$ , a cell signaling protein involved in systemic inflammation, and the abundance *Fusobacterium* level was reported (McCoy et al., 2013). On the other hand, the other two studies reported a lack of *Fusobacterium* enrichment in the adenoma group (Goedert et al., 2015; Xu & Jiang, 2017). Another study also reported no significant change in the abundance of *Fusobacterium* in the advanced and non-advanced adenoma group as compared to colonoscopically healthy control adults (Amitay et al., 2017). In our polyp cohort, *Fusobacterium* only showed a marginal decrease of its abundance in the biopsy, stool, and rectal swabs. It is possible that the pathways of polypogenesis are different from that of carcinogenesis and may depend on the invasiveness of the polyps which may, in turn, be reflected in observed discrepancies of association of *Fusobacterium* abundance. Additionally, the different experimental and analytical pipelines among the studies could be confounding this observation. A recent article that reviewed ninety original scientific published articles about *Fusobacterium* and its relationship with CRC in human, animal, and in vitro studies suggests that *F. nucleatum* can have a procarcinogenic effect for CRC which is probably is a result of triggering oncogenic and inflammatory pathways and alteration of the tumor immune environment (Hussan et al., 2017).

***Bacteroides***: *Bacteroides* species are butyrate-producing bacteria. Butyrate is a short chain fatty acid that has been shown to be effective in preventing inflammation through regulatory T-cells and controls proinflammatory cytokine expression (Cushing et al., 2015; Furusawa et al., 2013; Chang et al., 2014). This short chain fatty acid is also a histone deacetylase inhibitor and has a role in preventing colonic tumors and promoting normal cell proliferation, differentiation, and apoptosis. Additionally, butyrate modulates the Wnt signaling pathway involved in the development of colorectal cancer (Malcomson et al., 2015). Many *Bacteroides* species have been isolated from human stool. *Bacteroides fragilis* (*B. fragilis*) is the most common *Bacteroides* species found in clinical specimens, and it has been reported to have virulent *properties in some instances*. *Bacteroides* becomes a part of the intestinal microbiota early in life as it can pass from mother to the child during the vaginal birth (Reid, 2004). *B. fragilis* is generally considered to be a beneficial bacterium, but it can be pathogenic if it escapes from the gut to other body sites such as the abdomen, brain, liver, and lungs (Wexler, 2007).

Among the studies listed in Table S12, there are reports of both decrease and increase of this genus in the CRC and adenoma cohorts with respect to the control groups. The number of studies detecting a higher abundance of *Bacteroides* in the cancer states are higher than the number of studies observing lower abundance of this genus (Chen et al., 2012; Wu et al., 2013;

Zackular et al., 2014; Mira-Pascual et al., 2015; Thomas et al., 2016; Xu & Jiang et al., 2017; Gao et al., 2017; Yoon et al., 2017). However, for the adenoma cohorts, almost half detected enrichment and half reported depletion of *Bacteroides* in the adenoma state (Shen et al., 2010; Brim et al., 2013; Zackular et al., 2014; Yoon et al., 2017). There is even one study that reported both decrease and increase of *Bacteroides* taxa in the adenoma state which can be because of the differences in the subtaxa (Goedert et al., 2015). In our polyp dataset, all detected *Bacteroides* genera were higher in the polyp-Y group than polyp-N. As many different species and strains with potentially different functionalities can exist in the same genus, it is not that surprising that some of the *Bacteroides* taxa were decreased while others were increased in the disease state.

***Enterotoxigenic Bacteroides fragilis:*** *Enterotoxigenic B. fragilis* (ETBF) is a virulent bacterium that produces a toxin named fragilysin or B. fragilis toxin (BFT). BFT can stimulate both inflammatory responses and cell proliferation. The inflammatory effect of BFT occurs through activating the nuclear factor kappa B which stimulates inflammatory mediators. These mediators promote inflammation which is a risk factor for CRC (Sears et al., 2009; Shiryayev et al., 2013). The proliferation effect of ETBF occurs through activating the Wnt/  $\beta$ -catenin signaling pathway which increases cell proliferation (Sokol et al., 1999). In the APC minus mouse model, Wu et al., showed that ETBF could promote tumorigenesis and increases the chance of colon adenoma and tumor formation in mice colonized with ETBF compared to control mice (Wu et al., 2009). Tumorigenesis effect of ETBF can also occur through other pathways such as activating STAT3, inducing IL-17 cytokine, and inducing spermine oxidase-dependent reactive oxygen species (ROS) production which causes DNA damage (Tosolini et al., 2011; Goodwin et al., 2011). These studies suggest that there is a link between bacterial antigens, virulence factors, colon adenomas, and CRC. In our polyp dataset, the level of *B. fragilis* showed a reduction in the polyp-Y group. Thus, our study does not support *B. fragilis* playing a role in the polyposis pathway.

***Blautia:*** *Blautia* is a butyrate-producing bacterium that belongs to the *Firmicutes* phylum and Clostridial order. Most butyrate producers in the human colon belong to the *Firmicutes* phylum and in particular the clostridial clusters IV and XIVa (Louis and Flint, 2009; Van den Abbeele et al., 2013; Vital et al., 2014). *Blautia* is known to digest complex carbohydrates, and an abundance of these bacteria is a strong indication of a healthy gut. As such, it is usually reported that the abundance of *Blautia* decreases in colorectal cancer patients as compared to the healthy control individuals (Chen et al., 2012; Goedert et al., 2015; Xu & Jiang, 2017; Yoon et al., 2017). However, there are also reports of an increased abundance of this genus in the disease state (Wu et al., 2013; Mira-Pascual et al., 2015; Goedert et al., 2015; Yoon et al., 2017).

In the polyp dataset, two different OTUs were assigned as *Blautia* in the biopsy samples that were significantly lower in the polyp-Y group compared to the polyp-N group. In stool samples and rectal swabs, there were two OTUs classified as *Blautia*, and they were higher in the polyp-Y group. Based on these observations, *Blautia* was decreased in the CRC state, but in the adenoma state either increases or decreases can be seen. Differential abundance may depend on whether the adenoma is benign or aggressive suggesting that the potential effect of strains of this genus on the polyposis or carcinogenesis pathways may be different. As we did not have additional detailed clinical information about the patients with the polyps, we could not further clarify the reason behind this pattern in different sample types.

**Roseburia:** *Roseburia* spp. are commensal bacteria that produce short-chain fatty acids, particularly butyrate which affects colonic motility and immunity maintenance and has anti-inflammatory properties. Depletion in *Roseburia* spp. abundance may affect various metabolic pathways and be associated with several diseases including irritable bowel syndrome, obesity, type 2 diabetes, nervous system conditions, and allergies. *Roseburia* spp. could also serve as probiotics for the restoration of a beneficial microbiota (Tamanai-Shacoori et al., 2016). As *Roseburia* produces large amounts of butyrate by fermenting dietary carbohydrates, it may be critical for the control of inflammatory processes, especially in the colon (Louis et al., 2010; Louis et al., 2014; Pryde et al., 2002; Tamanai-Shacoori et al., 2016).

Two of the seventeen studies reported enrichment of *Roseburia* in CRC samples, and two other studies showed depletion of this taxon in the cancer group (Marchesi et al., 2011; Thomas et al., 2016; Chen et al., 2012; Wu et al., 2013). None of the adenoma studies we investigated showed changes in *Roseburia* taxa. In our polyp dataset, we did detect a decrease in the abundance of this taxon in biopsy samples and an increase in abundance in the swab and fecal samples of the polyp-Y group, although these changes were not statistically significant.

**Faecalibacterium:** *Faecalibacterium* is commonly present in the gastrointestinal tract and is recognized as a commensal bacterium. *Faecalibacterium prausnitzii* (*F. prausnitzii*) is a dominant species of the Clostridium leptum group and is one of the most abundant anaerobic bacteria in the human gut (Arumugam et al., 2011). *F. prausnitzii* plays a key role in maintaining intestinal health and providing energy to the colonocytes (Louis & Flint, 2009). It has also been shown that *F. prausnitzii* levels were decreased in IBD patients compared with healthy control controls (Yang et al., 2008). Three of the studies listed in Table S12 found that *Faecalibacterium* spp. were increased in adenoma subjects (Shen et al., 2010; Brim et al., 2013; Yoon et al., 2017) and Marchesi et al. reported an increase of this genus in CRC cases (Marchesi et al., 2011). Chen et al. reported an increase of *Faecalibacterium* in CRC tissue while they found a decrease of this taxon in the swab samples (Chen et al., 2012). Four other studies demonstrated a decrease of *Faecalibacterium* in the CRC group (Wu et al., 2013; Xu & Jiang, 2017; Gao et al., 2017; Yoon et al., 2017). In our polyp dataset, biopsies and stool samples had a higher abundance of *F. prausnitzii*. However, the swabs showed a lower abundance of this taxon. Thus, we did not find a consistent pattern for the change of *F. prausnitzii* with respect to adenoma and CRC based on these studies.