

Supplementary information

Extended Data Table 1. OTUs not in the database

OTU ID	Sequence
OTU65	TGGGGAATATTGCACAATGGGGGAAACCCTGATGCAGCAACGCCG CGTGAAGGATGACGGTTTTTCGGATTGTAAACTTCTTTTCTTAGTGA CGAAGACAGTGACGGTAGCTAAGGAATAAGCATCGGCTAACTACG TGCCAGCAGCCGCGGTAAATACGTAGGATGCAAGCGTTATCCGGAT TTACTGGGTGTAAAGGGAGCGCAGGCGGGACTGCAAGTTGGATGT GAAATACCGTGGCTTAACCACGGAAGTGCATCCAAAAGTGTAGTT CTTGAGTGAAGTAGAGGCAAGCGGAATTCCGAGTGTAGCGGTGAA ATGCGTAGATATTCGGAGGAACACCAGTGGCGAAGGCGGCTTGCT GGGCTTTAACTGACGCTGAGGCTCGAAAGTGTGGGGAGCAAACAG G
OTU192	TGGGGAATATTGGGCAATGGGGGGAACCCTGACCCAGCAACGCCG CGTGGAGGAAGAAGGTTTTTCGGATCGTAAACTCCTGTCCTTGGAG ACGAGTAGAAGACGGTATCCAAGGAGGAAGCCCCGGCTAACTACG TGCCAGCAGCCGCGGTAAATACGTAGGGGGCAAGCGTTGTCCGGAA TAATTGGGCGTAAAGGGCGCGTAGGCGGCTCGGTAAGTCTGGAGT GAAAGTCCTGCTTTTAAGGTGGGAATTGCTTTGGATACTGTCCGGC TTGAGTGCAGGAGAGGTTAGTGGGAATTCCCAGTGTAGCGGTGAAA TGCGTAGAGATTGGGAGGAACACCAGTGGCGAAGGCGACTAACTG GACTGTAAGTACGCTGAGGCGCGAAAGTGTGGGGAGCAAACAGG TGGGGAATATTGCACAATGGGGGAAACCCTGATGCAGCGACGCCG CGTGAGTGAAGAAGTATTTTCGGTATGTAAAGCTCTATCAGCAGGG AAGAGAATGACGGTACCTGACTAAGAAGCCCCGGCTAACTACGTG CCAGCAGCCGCGGTAAATACGTAGGGGGCAAGCGTTATCCGGATTT
OTU163	ACTGGGTGTAAAGGGAGCGTAGACGGCTTAGCAAGTCTGAAGTGA AAGCCCCGGGGCTCAACCCCCGGGACTGCTTTGGAACTGTTAAGCT GGAGTGCTGGAGAGGTAAGCGGAATTCCTAGTGTAGCGGTGAAAT GCGTAGATATTAGGAGGAACACCAGTGGCGAAGGCGGCTTACTGG ACAGTAACTGACGTTGAGGCTCGAAAGCGTGGGGAGCAAACAGG TGGGGAATATTGCACAATGGAGGAAACTCTGATGCAGCGACGCCG CGTGAGTGAAGAAGTAGTTCGCTATGTAAAGCTCTATCAGCAGGG AAGATAGTGACGGTACCTGACTAAGAAGCTCCGGCTAAATACGTG CCAGCAGCCGCGGTAAATACGTATGGAGCAAGCGTTATCCGGATTT
OTU14	ACTGGGTGTAAAGGGAGTGTAGGTGGCCAGGCAAGTCAGAAGTGA AAGCCCCGGGGCTCAACCCCCGGGACTGCTTTTGAACTGCAGGGCT AGAGTGCAGGAGGGGCAAGTGGGAATTCCTAGTGTAGCGGTGAAAT GCGTAGATATTAGGAGGAACACCAGTGGCGAAGGCGGCTTGCTGG ACTGTAAGTACACTGAGGCTCGAAAGCGTGGGGAGCAAACAGG

OTU487 TGGGGAATATTGCACAATGGGCGAAAGCCTGATGCAGCAACGCCG
CGTGAAGGAAGAAGGTCTTCGGATCGTAAACTTCTGTCCTTGGGG
AAGATAATGACGGTACCCTTGGAGGAAGCCCCGGCTAACTACGTG
CCAGCAGCCGCGGTAATACGTAGGGGGCAAGCGTTATCCGGAATT
ATTGGGCGTAAAGAGTGCGTAGGTGGTTACCTAAGCAGGGGGTGA
AAGGCACTGGCTTAACCAATGTCAGCCCCCTGAACTGGGTACCTTG
AGTGCAGGAGAGGAAAGCGGAATTCCTAGTGTAGCGGTGAAATGC
GTAGATATTAGGAGGAACACCAGTGGCGAAGGCGGCTTTCTGGAC
TAGGGAATATTGGGCAATGGGCGCAAGCCTGACCCAGCAACGCCG
CGTGAGGGAAGAAGGTTTTTCGGATTGTAAACCTCTGTCCTATGTGA
CGAAGGAAGTGACGGTAGCATAGGAGGAAGCCCCGGCTAACTACG
TGCCAGCAGCCGCGGTAATACGTAGGGGGCGAGCGTTGTCCGGAA
OTU172 TTAAGTGGGCGTAAAGGGTGCCTAGGCGGTTTGGTAAGTTGGATGT
GAAATACCCGGGCTTAAGTGGGGGCTGCATCCAATACTGTCGGA
CTTGAGTGCAGGAGAGGAAAGCGGAATTCCTAGTGTAGCGGTGAA
ATGCGTAGATATTAGGAGGAACACCGGTGGCGAAGGCGGCTTTCT
GGACTGTAAGTACGCTGAGGCACGAAAGCGTGGGGAGCAAACA
GG
TGAGGAATATTGGTCAATGGACGAGAGTCTGAACCAGCCAAGTAG
CGTGAAGGATGACTGCCCTATGGGTGTAAACTTCTTTTATATGGG
AATAAAAAAGGCCACGTGTGGCCTATTGTATGTACCTTATGAATAA
GCATCGGCTAATTCCGTGCCAGCAGCCGCGGTAATACGGAAGATG
OTU694 CGAGCGTTATCCGGATTTATTGGGTTTAAAGGGAGCGTAGGCGGG
CGATTAAGTCAGCGGTAAAATAGTGTGGCTCAACCATGCTCTGCCG
TTGATACTGGTTGCCTTGAGTGCACACAAGGAAGATGGAATTCGTG
GTGTAGCGGTGAAATGCTTAGATATCACGAAGAACTCCGATTGCG
AAGGCAGTCTTCTGGGGTGTTACTGACGCTGAGGCTCGAAAGTGC
GGGTATCAAACAGG
TGGGGAATATTGCACAATGGGCGAAAGCCTGATGCAGCAACGCCG
CGTGAGGGAAGACGGTTTTTCGGATTGTAAACCTCTGTTTTTCGGTGA
CGAACAAATGACGGTAACCGAGTAGGAAGCCACGGCTAACTACGT
GCCAGCAGCCGCGGTAATACGTAGGTGGCAAGCGTTATCCGGAAT
OTU132 TACTGGGTGTAAAGGGAGCGCAGGCGGGATAGCAAGTCAGCTGTG
AAAATATGGGCTCAACCCATAAACTGCAGTTGAAACTGTTATTCT
TGAGTGGAGTAGAGGCAAGCGGAATTCGAGTGTAGCGGTGAAAT
GCGTAGATATTCGGAGGAACACCAGTGGCGAAGGCGGCTTGCTGG
GCTCTAACTGACGCTGAGGCTCGAAAGTGTGGGGAGCAAACAGG

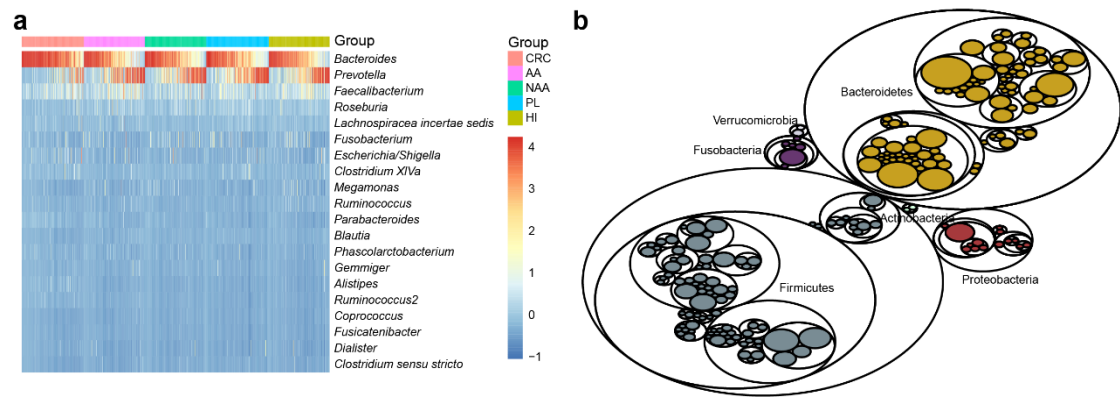
Extended Data Table 2. Comparison of pathway

Literature report	This research
Cys and Met metabolism	L-Methionine Biosynthesis III
Central carbohydrate metabolism; Other carbohydrate metabolism	Sucrose Degradation IV (Sucrose Phosphorylase); Pentose Phosphate Pathway (Non-Oxidative Branch); Glycolysis I (From Glucose 6-Phosphate)
Arg biosynthesis; Arg and Pro metabolism	Arginine, Ornithine and Proline Interconversion
Cofactor and vitamin biosynthesis	Reductive Acetyl Coenzyme A Pathway; Methylmalonyl-CoA Pathway
Aromatic amino acid metabolism	1,4-Dihydroxy-6-Naphthoate Biosynthesis I
Lys biosynthesis; Lys degradation	L-Lysine Fermentation to Acetate and Butanoate
Methane metabolism	Methanogenesis from Acetate
Ala, Asp and Glu metabolism	Methylaspartate Cycle
NA	Pyridoxal 5'-Phosphate Biosynthesis I
NA	Super Pathway of (R,R)-Butanediol Biosynthesis
NA	Bifidobacterium Shunt
NA	Chorismate Biosynthesis II (Archaea)
NA	Urea Cycle
NA	Superpathway of Adenosine Nucleotides <i>de novo</i> Biosynthesis I
NA	Adenosine Deoxyribonucleotides <i>de novo</i> Biosynthesis II
NA	Guanosine Ribonucleotides <i>de novo</i> Biosynthesis
NA	Acetylene Degradation
NA	Cmp-Legionaminate Biosynthesis I
NA	Stearate Biosynthesis II (Bacteria And Plants)
NA	Pyrimidine Deoxyribonucleotides <i>de novo</i> Biosynthesis IV
NA	Lipid Iva Biosynthesis
NA	O-Antigen Building Blocks Biosynthesis (<i>E. coli</i>)
NA	Palmitate Biosynthesis II (Bacteria And Plants)
NA	Isopropanol Biosynthesis
NA	Super Pathway of Pyrimidine Deoxyribonucleoside Salvage
NA	Chondroitin Sulfate Degradation I (Bacterial)

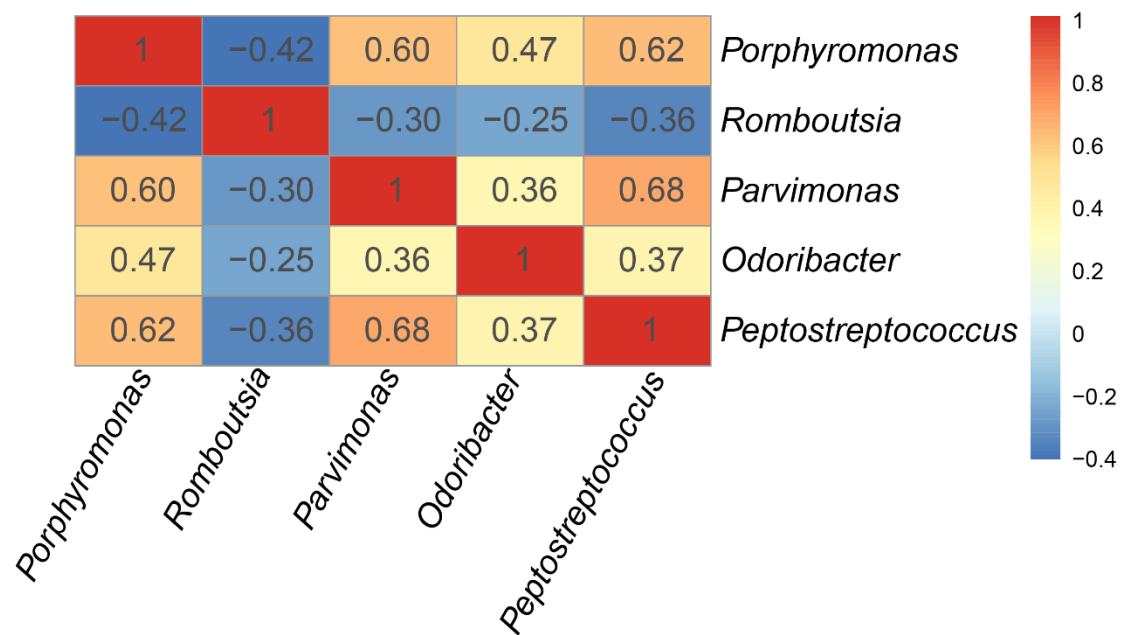
Extended Data Table 3. Importance of genera and OTUs in the relative OTU abundances model

Genera (Rank/Total)	OTUs in the genus	OTU-model	Rank in the OTUs model	Importance
<i>Alistipes</i> (7/315)	7	OTU252	9	0.0052
		OTU51	33	0.0035
<i>Anaerostipes</i> (36/315)	4	OTU382	32	0.0035
		OTU135	43	0.0033
		OTU138	23	0.0039
		OTU153	3	0.0088
		OTU164	49	0.0031
<i>Bacteroides</i> (46/315)	102	OTU17	53	0.0030
		OTU298	36	0.0034
		OTU3559	37	0.0034
		OTU43	24	0.0039
		OTU520	15	0.0047
		OTU59	44	0.0032
		OTU325	1	0.0127
		OTU361	41	0.0033
		OTU44	13	0.0049
<i>Clostridium sensu stricto</i> (15/315)	6	OTU106	11	0.0051
		OTU186	6	0.0061
		OTU367	4	0.0071
<i>Clostridium XIVa</i> (40/315)	15	OTU227	34	0.0035
		OTU496	21	0.0041
<i>Clostridium XIVb</i> (24/315)	7	OTU188	20	0.0041
		OTU501	46	0.0032
<i>Dialister</i> (37/315)	4	OTU511	2	0.0126
<i>Faecalibacterium</i> (50/315)	7	OTU6392	52	0.0030
<i>Fusobacterium</i> (70/315)	7	OTU120	12	0.0050
		OTU527	14	0.0048
<i>Lachnospiracea incertae sedis</i> (19/315)	10	OTU54	42	0.0033
<i>Parabacteroides</i> (53/315)	16	OTU299	19	0.0042
<i>Prevotella</i> (28/315)	61	OTU1	51	0.0030
		OTU108	8	0.0054
		OTU123	26	0.0038
		OTU18	10	0.0051
		OTU189	48	0.0032
		OTU197	17	0.0046

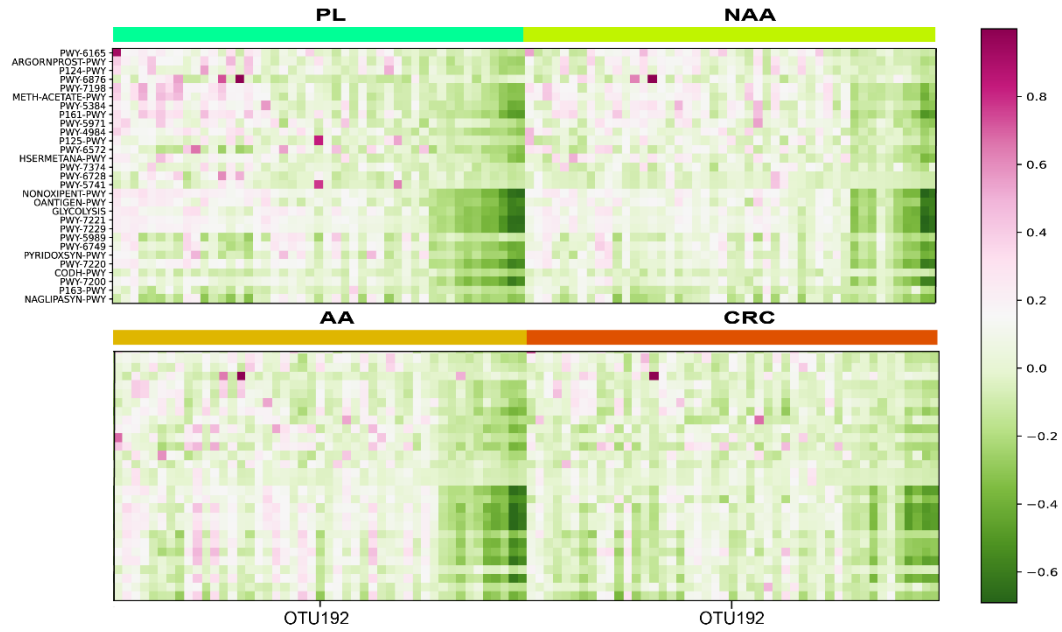
		OTU202	5	0.0063
		OTU29	18	0.0042
		OTU31	28	0.0037
		OTU333	35	0.0034
		OTU34	7	0.0055
		OTU50	38	0.0034
<i>Romboutsia</i> (1/315)	1	OTU87	25	0.0038
<i>Roseburia</i> (10/315)	5	OTU25	22	0.0040
<i>Ruminococcus2</i> (29/315)	4	OTU53	16	0.0046
		OTU132	50	0.0031
		OTU14	31	0.0036
		OTU163	30	0.0036
		OTU172	40	0.0033
Unclassified (NA)	158	OTU182	47	0.0032
		OTU192	29	0.0037
		OTU487	39	0.0034
		OTU65	27	0.0037
		OTU694	45	0.0032



Extended Data Fig. 1 Alterations in gut microbial composition with CRC progression. **a** The heat map of the relative abundances of OTUs with CRC progression. Group represents the five stages of CRC development (Healthy Individual-HI; Polyps-PL; Non-Advanced Adenoma -NAA; Advanced Adenoma-AA; Colorectal Cancer-CRC). The intensity of colors in the heat map indicates the relative abundance of the corresponding OTU. **b** Important OTUs annotated from metagenome at different taxonomic levels. The outmost circles represent phylum, and the inner circles represent class, family, and genus, respectively. The circle size represents the mean relative abundance of taxa, the colors indicate phylum.



Extended Data Fig. 2 Correlation between genus biomarkers in the random forest classification model using the relative genus abundances. Shown is the Spearman correlation coefficient between the top 5 genera in the model. The numbers indicate the correlation between the corresponding genera.



Extended Data Fig. 3 Predicted microbial functional changes in serial CRC stages. The horizontal axis represents the 54 important OTUs at each stage of colorectal disease, and the vertical axis is the most important signal pathways predicted with PICRUSt2. Colors indicate the Spearman correlation coefficient between the OTUs and signal pathways.