
Noninvasive, microbiome-based diagnosis of inflammatory bowel disease

In the format provided by the
authors and unedited

Table S1 Characteristics of study participants in discovery cohort.

	CD (N=174)	UC (N=205)	Controls (N=118)
Age			
Mean (SD)	42.8 (14.5)	53.3 (13.3)	42.7 (6.90)
Gender (N / percentage)			
F	62 (35.6%)	105 (51.2%)	75 (63.6%)
Disease location (L) (N / percentage)			
L1 (terminal ileum)	38 (21.8%)		
L2 (colon)	53 (30.5%)		
L3 (ileocolon)	83 (47.7%)		
L4 (upper GI)	15 (8.6%)		
Disease behavior (B) (N / percentage)			
B1, non-stricturing, non-penetrating	104 (59.8%)		
B2, stricturing	40 (23.0%)		
B3, penetrating	43 (22.4%)		
Disease extent (E) (N / percentage)			
E1 (proctitis)		67 (32.7%)	
E2 (left-sided colitis)		68 (33.2%)	
E3 (extensive colitis)		70 (34.1%)	
Medication (N / percentage)			
5-ASA/Mesalazine	161 (92.5%)	200 (97.6%)	
Steroids	136 (78.2%)	110 (53.7%)	
Immunosuppressant	130 (74.7%)	57 (27.8%)	
Biologics	23 (13.2%)	0 (0%)	

Table S2 Characteristics of the cohorts in this study.

Dataset	Group	Sample size	Age (average ± s.d.)	Gender F (%)	Country/R egion	Sequencing platform	Accession No.
Discovery Cohort	Control	118	42.73 ± 6.90	63.6	Hong Kong, China	Illumina NovaSeq 6000	
	UC	205	53.30 ± 13.30	51.2			
	CD	174	42.80 ± 14.50	35.6			
Validation Cohort	Control	139	54.28 ± 11.74	46.8	Hong Kong, China	Illumina NovaSeq 6000	
	UC	139	55.28 ± 13.06	44.6			
	Control	108	48.10 ± 8.89	57.1	Hong Kong, China	Illumina NovaSeq 6000	
	CD	92	39.4 ± 13.30	52.0			
	Control	81	47.6 ± 13.40	53.1	Australia	Illumina NovaSeq 6000	
	CD	98	38.30 ± 13.70	52.0			
Public datasets	Control	34	34.47 ± 15.04	NA	the United States	Illumina HiSeq 2000	PRJNA400072
	UC	53	45.53 ± 19.28	NA			
	CD	68	42.31 ± 14.98	NA			
	Control	22	40.82 ± 8.90	NA	Netherland s	Illumina HiSeq 2000	PRJNA400072
	UC	23	50.00 ± 21.89	NA			
	CD	20	45.15 ± 11.50	NA			
	Control	15	NA	NA	Mainland China	llumina HiSeq X Ten	PRJNA429990
	UC	25	NA	NA			
	Control	54	20.70 ± 7.76	5.6	Mainland China	Illumina HiSeq 2000	PRJEB15371
	CD	48	28.53 ± 7.89	22.9			
	Control	1040	43.27 ± 11.66	58.8	Netherland s	Illumina HiSeq	EGAS00001001704 EGAD00001004194
	UC	119	NA	50.4			
	CD	216	NA	72.7			
	Control	74	NA	NA	the United States	Illumina MiSeq	PRJNA385949
	UC	97	NA	NA			
	CD	88	NA	NA			
	Control	198	42.63 ± 13.84	32.3	the United States	Illumina HiSeq	PRJNA398089
	UC	236	35.38 ± 9.34	82.2			
	CD	358	36.93 ± 12.16	62.6			
	Control	224	NA	54.0	Spain Denmark	Illumina HiSeq	PRJEB1220
	UC	125	NA	61.6			
	CD	21	NA	66.7			
	Control	38	32.00 ± 16.16	NA	United Kingdom	Illumina HiSeq	PRJEB18780
	CD	56	11.43 ± 2.30	NA			
Total	Control	2145	4406				
	UC	1022					
	CD	1239					

Table S3 Sample size of non-IBD disease cohorts.

Study	Sequencing platform	Country/Region	CRC	CA	IBS	Obesity	CVD	Controls	Total
Qi S	Illumina NovaSeq	Hong Kong, China	160	162	117	148	143		730
Feng Q	Illumina HiSeq	Austria	10	4				6	20
Zeller G	Illumina HiSeq	France	25	25				41	91
Wirbel J	Illumina HiSeq	Germany	33					53	86
Yachida S	Illumina HiSeq	Japan	144	39				136	319
Mars RAT	Illumina NextSeq	the United States			29			24	53
Le Chatelier E	Illumina HiSeq	Denmark				170		104	274
		Total	372	230	146	318	143	364	1573

CRC	Colorectal cancer
CA	Colorectal adenomas
IBS	Irritable bowel syndrome
CVD	Cardiovascular disease

Table S4 Primers and probes used in multiplex droplet digital PCR.**Ulcerative colitis**

	Bacterial Species	Primers	Sequence (5'-->3')	Concentration (nM)
Reaction 1	<i>Blautia hansenii</i>	F	ACGTGTAGAGTGGGCAAAAA	900
		R	CATTTTCCTCTCCTTGCGGA	
	<i>Clostridium leptum</i>	F	GGACTGAGCTTTCCAACCTT	
		R	CATGTCATTGAGCCAGCCTA	
	<i>Actinomyces sp. oral taxon 181</i>	F	CGTGTGGAGTGCTATTGTC	
		R	TATTCATCCTCCTCAGGCGT	
	<i>Gemella morbillorum</i>	F	GACTCTAACCGTGCGCTAAT	
		R	GATGCTCCATTCTGTTCCA	
Reaction 2	<i>Clostridium spiroforme</i>	F	CAAAGCTATTGCAGTTGGACA	
		R	TTGCTCATGCCAGAACGATT	
	<i>Fusicatenibacter saccharivorans</i>	F	TGGTTGCACGAATACCCATT	
		R	CTCCTGCATAATGAGTGCCA	
	<i>Odoribacter splanchnicus</i>	F	GACAATCAACTCGACGGGAA	
		R	TTTTGTACGTCCTCAGGCAG	
	<i>Gemmiger formicilis</i>	F	GGAGAAGAGGGAAAGCAGTC	
		R	TGGTACAAAACACATCGCCA	
Reaction 3	<i>Ruminococcus torques</i>	F	CATATGAGGCTCTGCTGCTT	
		R	GGAACGTACTTCTCTTCCG	
	<i>Bilophila wadsworthia</i>	F	AGGAAGCCGTCGAAAAGTAC	
		R	AGACGGCAAGTTCTTCTTCC	
	16s reference gene	F	CGTCAGCTCGTGYCGTGAG	
		R	CGTCRTCCCCRCCTTCC	

	Bacterial Species	Fluorophore-Quencher	Sequence (5'-->3')	Concentration (nM)
Reaction 1	<i>Blautia hansenii</i>	FAM-BHQ1	CCGGTGACGCGCAGGGAGCA	180
	<i>Clostridium leptum</i>	FAM-BHQ1	CCCACCGATTGCCCAGCGC	400
	<i>Actinomyces sp. oral taxon 181</i>	HEX-MGB	TCCGCTCCTGAGTCAGTGGGTTG	180
	<i>Gemella morbillorum</i>	HEX-MGB	CGAACATGCAACGTCAAGCA	450
Reaction 2	<i>Clostridium spiroforme</i>	FAM-BHQ1	ACGCTCTCCACCATGACCAAACGGAGC	120
	<i>Fusicatenibacter saccharivorans</i>	FAM-BHQ1	TGCAAACGGTCTGCCAGATCCGAACCA	480
	<i>Odoribacter splanchnicus</i>	HEX-BHQ1	CCGAAACCGAGTTGACCCACACCAGCC	200
	<i>Gemmiger formicilis</i>	HEX-BHQ1	ACGGTTGCTTCCATCCAGAGATGCGGC	450
Reaction 3	<i>Ruminococcus torques</i>	HEX-BHQ1	GGTTCTGGAGGAGAGGGATCCGGGCG	480
	<i>Bilophila wadsworthia</i>	HEX-MGB	CGCGACTCCTGAAACCCTG	120
	16s reference gene	FAM-BHQ1	TTAAGTCCCRYAACGAGCGCAACCC	200

Crohn's disease

	Bacterial Species	Primers	Sequence (5'-->3')	Concentration (nM)
Reaction 1	<i>Bacteroides fragilis</i>	F	TCTGGTTATCCGGACTCCTT	900
		R	TCTTTTTCTGCCGGTACCTG	
	<i>Eubacterium sp. CAG: 274</i>	F	TCATTTGTTTGGGTGCTGTC	
		R	TCAAAGCCCATATAGCGAAGT	
	<i>Escherichia coli</i>	F	AATGACTGGTGCATTCCGAA	
		R	AACGCTTTGGCTAAGATCGA	
	<i>Actinomyces sp. oral taxon 181</i>	F	CGTGTTGGAGTGCTATTTGC	
		R	TATTCATCCTCCTCAGGCGT	
Reaction 2	<i>Roseburia inulinivorans</i>	F	GGAACCAATGGAAAACGTGTG	
		R	AACCCCTTCCCATATCCATAACG	
	<i>Ruminococcus obeum</i>	F	CACCTGGATACCGTAAAGAAC	
		R	GAAGCCCGACTGTTTGAG	
	<i>Roseburia intestinalis</i>	F	TCGGATGGATCGTCTATCTG	
		R	CATCCTTGAAAAGACCGTGAG	
	<i>Lawsonibacter asaccharolyticus</i>	F	TTGACGCCTTGACCATATCC	
		R	TAAAGTCCTGAAACTGCCGG	
Reaction 3	<i>Dorea formicigenerans</i>	F	CAGAGAATGGCAGACGGAAA	
		R	CTTCATAGTTGCCTCCGG	
	16s reference gene	F	CGTCAGCTCGTGYCGTGAG	
		R	CGTCRTCCCCRCCTTCC	

	Bacterial Species	Fluorophore-Quencher	Sequence (5'-->3')	Concentration (nM)
Reaction 1	<i>Bacteroides fragilis</i>	FAM-BHQ1	ACCTGCAACGGCACCATAGCGGC	120
	<i>Eubacterium sp. CAG: 274</i>	FAM-BHQ1	AGGGTTCTGCCCCTCAGCCTTGGC	400
	<i>Escherichia coli</i>	HEX-BHQ1	TGGCGAAATCAGACAGCCCAATGCGACG	120
	<i>Actinomyces sp. oral taxon 181</i>	HEX-MGB	TCCGCTCCTGAGTCAGTGGGTTG	400
Reaction 2	<i>Roseburia inulinivorans</i>	FAM-BHQ1	TGCCAAACACGCTTCCCGAACAATGA	170
	<i>Ruminococcus obeum</i>	FAM-MGB	AATATGATGGCGCACTGGGC	480
	<i>Roseburia intestinalis</i>	HEX-BHQ1	TGCCTGCAGTTTGACACCGGAACCG	180
	<i>Lawsonibacter asaccharolyticus</i>	HEX-BHQ1	AGCGGAAGCTGGCGCAGTAGGTGC	450
Reaction 3	<i>Dorea formicigenerans</i>	HEX-BHQ1	CGGAGCAGGCTGCGGACCGG	250
	16s reference gene	FAM-BHQ1	TTAAGTCCCRYAACGAGCGCAACCC	350