

**Supplementary Table S1. Characteristics of recruited subjects in each group**

	Advanced adenoma		Nonadvanced adenoma		Healthy controls		P-value		
	N=268	%	N=490	%	N=788	%	P <sup>a</sup>	P <sup>b</sup>	P <sup>c</sup>
<b>Gender</b>							<0.001	<0.001	<0.001
Male	197	73.5	317	64.7	378	48.0			
Female	71	26.5	173	35.3	410	52.0			
<b>Age (years)</b>							<0.001	0.002	<0.001
50-54	44	16.4	84	17.1	202	25.6			
55-59	53	19.8	98	20.0	170	21.6			
60-64	67	25.0	127	25.9	201	25.5			
65-69	78	29.1	130	26.5	168	21.3			
70-74	26	9.7	51	10.4	47	6.0			
<b>BMI (Chinese categories, kg/m<sup>2</sup>)</b>							0.003	0.001	0.030
<24 (Underweight & normal)	110	41.0	224	45.7	411	52.2			
24-27.9 (Overweight)	124	46.3	217	44.3	323	41.0			
≥28 (Obese)	34	12.7	49	10.0	54	6.9			
<b>Family history</b>							0.424	0.277	0.986
Yes	38	14.2	55	11.2	90	11.4			
No	230	85.8	435	88.8	698	88.6			
<b>Smoking</b>							<0.001	<0.001	0.001
Current smoker	100	37.3	149	30.4	166	21.0			
Past smoker	16	6.0	29	5.9	55	7.0			
Non-smoker	152	56.7	312	63.7	567	72.0			
<b>Alcohol use</b>							0.001	<0.001	0.321

Regular	87	32.5	118	24.1	165	20.9			
Seldom	46	17.2	77	15.7	117	14.8			
No	135	50.4	295	60.2	506	64.2			
Physical activity, MET-h/week							0.004	0.001	0.042
<31.65	67	25.0	127	25.9	192	24.4			
31.65-69.4	75	28.0	128	26.1	185	23.5			
69.5-124.2	80	29.9	129	26.3	184	23.4			
>=124.3	46	17.2	106	21.6	227	28.8			
FIT <sup>a</sup>							<0.001	<0.001	1
+	26	10.5	8	1.8	12	1.6			
-	222	89.5	444	98.2	720	98.4			
APCS score <sup>a</sup>							<0.001	<0.001	<0.001
0	5	2.0	12	2.7	51	7.0			
1	19	7.7	46	10.2	154	21.0			
2	43	17.3	92	20.4	171	23.4			
3	58	23.4	119	26.3	148	20.2			
4	86	34.7	144	31.9	144	19.7			
5	33	13.3	36	8.0	55	7.5			
6	4	1.6	3	0.7	9	1.2			

Abbreviations: BMI, body mass index; MET, metabolic equivalent; FIT, fecal immunochemical test; APCS, Asia-Pacific Colorectal Screening.

<sup>a</sup> Total number of participants equals to 248, 452 and 732 for advanced adenoma group, nonadvanced adenoma group and healthy control group, respectively, due to some with no available FIT results.

**Supplementary Table S2. Aggregation of features identified using LEfSe and MaAsLin methods.**

<b>Feature</b>	<b>Enrichment status</b>	<b>Methods</b>
k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales	AA	LEfSe
k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Actinomycetaceae	AA	LEfSe
k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Actinomycetaceae;g__Actinomyces	AA	LEfSe
k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Actinomycetaceae;g__Actinomyces	AA	LEfSe
;s__k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Bifidobacteriales;f__Bifidobacteriaceae;g__Bifidobacterium;s__Bifidobacterium bifidum	HC	LEfSe
k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Micrococcales;f__Micrococcaceae;g__Rothia;s__uncultured bacterium	AA	MaAsLin
k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Bacteroidaceae;g__Bacteroides;s__Bacteroides thetaiotaomicron	HC	LEfSe
k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Marinifilaceae	HC	LEfSe
k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Marinifilaceae;g__Odoribacter	HC	LEfSe
k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Marinifilaceae;g__Odoribacter;s__gut metagenome	HC	LEfSe
k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales	HC	LEfSe
k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Family XI	AA	LEfSe + MaAsLin
k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Family XI;g__Gemella	AA	LEfSe + MaAsLin
k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Family XI;g__Gemella;s__uncultured bacterium	AA	LEfSe + MaAsLin
k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae	AA	MaAsLin
k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus	AA	MaAsLin
k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;s__	AA	LEfSe + MaAsLin
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiaceae 1;g__Clostridium sensu stricto 1	AA	MaAsLin

k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiaceae 1;g__Clostridium sensu stricto 1;_s	AA	MaAsLin
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiaceae 1;g__Clostridium sensu stricto 1;s__uncultured bacterium	AA	LEfSe + MaAsLin
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Blautia;s__Blautia massiliensis	HC	LEfSe
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Lachnospira;_	HC	LEfSe
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Lachnospiraceae UCG-004	HC	LEfSe
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Lachnospiraceae UCG-004;_	HC	LEfSe
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Tyzzerella 4	AA	LEfSe + MaAsLin
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Tyzzerella 4;s__	AA	LEfSe + MaAsLin
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Peptostreptococcaceae	AA	LEfSe
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Peptostreptococcaceae;g__Terrisporobacter	AA	LEfSe
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Peptostreptococcaceae;g__Terrisporobacter;s__uncultured bacterium	AA	LEfSe
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae	HC	LEfSe
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;_	HC	LEfSe
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;_;	HC	LEfSe
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__Faecalibacterium	HC	LEfSe
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__Faecalibacterium;_	HC	LEfSe
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__Ruminococcaceae NK4A214 group;s__gut metagenome	HC	LEfSe
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__Ruminococcaceae UCG-002	HC	LEfSe
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__Ruminococcaceae UCG-002;_	HC	LEfSe
k__Bacteria;p__Firmicutes;c__Negativicutes;o__Acidaminococcales	AA	LEfSe
k__Bacteria;p__Firmicutes;c__Negativicutes;o__Acidaminococcales;f__Acidaminococcaceae	AA	LEfSe

k__Bacteria;p__Firmicutes;c__Negativicutes;o__Acidaminococcales;f__Acidaminococcaceae;g__Phascolarctobacterium	AA	LEfSe
k__Bacteria;p__Fusobacteria	AA	LEfSe + MaAsLin
k__Bacteria;p__Fusobacteria;c__Fusobacteriia	AA	LEfSe + MaAsLin
k__Bacteria;p__Fusobacteria;c__Fusobacteriia;o__Fusobacteriales	AA	LEfSe + MaAsLin
k__Bacteria;p__Fusobacteria;c__Fusobacteriia;o__Fusobacteriales;f__Fusobacteriaceae	AA	LEfSe + MaAsLin
k__Bacteria;p__Fusobacteria;c__Fusobacteriia;o__Fusobacteriales;f__Fusobacteriaceae;g__Fusobacterium	AA	LEfSe
k__Bacteria;p__Fusobacteria;c__Fusobacteriia;o__Fusobacteriales;f__Fusobacteriaceae;g__Fusobacterium;s__	AA	MaAsLin
k__Bacteria;p__Fusobacteria;c__Fusobacteriia;o__Fusobacteriales;f__Fusobacteriaceae;g__Fusobacterium;s__uncultured organism	AA	LEfSe
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Betaproteobacteriales	HC	LEfSe
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Betaproteobacteriales;f__Burkholderiaceae	HC	LEfSe

Abbreviations:

k\_, kingdom; p\_, phylum, c\_, class; o\_, order; f\_, family; g\_ genera; s\_, species.

AA, advanced adenoma; HC, healthy control.

**Supplementary Table S3. Diagnostic performance of 13 fecal microbial markers in discriminating advanced adenoma and healthy control using the least absolute shrinkage and selection operator (LASSO) binary logistic regression model.**

Feature	Apparent AUC	0.632+ AUC	Enrichment status	Methods
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Tyzzerella 4	0.578 (0.542-0.614)	0.545 (0.52-0.610)	AA	MaAsLin + LEfSe
k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Family XI;g__Gemella	0.567 (0.535-0.599)	0.535 (0.52-0.602)	AA	MaAsLin + LEfSe

k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__Faecalibacterium;__	0.551 (0.511-0.591)	0.531 (0.498-0.593)	HC	LefSe
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Lachnospiraceae UCG-004	0.546 (0.518-0.574)	0.527 (0.509-0.579)	HC	LefSe
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;__	0.542 (0.51-0.573)	0.524 (0.497-0.580)	HC	LefSe
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiaceae 1;g__Clostridium sensu stricto 1	0.538 (0.498-0.579)	0.522 (0.485-0.586)	AA	MaAsLin
k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus	0.525 (0.484-0.566)	0.521 (0.481-0.576)	AA	MaAsLin
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiaceae 1;g__Clostridium sensu stricto 1;s__uncultured bacterium	0.541 (0.502-0.58)	0.52 (0.485-0.581)	AA	MaAsLin
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Lachnospira;__	0.541 (0.502-0.580)	0.514 (0.488-0.580)	HC	LefSe
k__Bacteria;p__Fusobacteria;c__Fusobacteriia;o__Fusobacteriales;f__Fusobacteriaceae;g__Fusobacterium;s__uncultured organism	0.54 (0.511-0.570)	0.513 (0.497-0.564)	AA	LefSe
k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Micrococcales;f__Micrococcaceae;g__Rothia;s__uncultured bacterium	0.523 (0.492-0.555)	0.508 (0.481-0.559)	AA	MaAsLin
k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Bifidobacteriales;f__Bifidobacteriaceae;g__Bifidobacterium;s__Bifidobacterium bifidum	0.524 (0.509-0.539)	0.504 (0.504-0.540)	HC	LefSe
k__Bacteria;p__Fusobacteria;c__Fusobacteriia;o__Fusobacteriales;f__Fusobacteriaceae;g__Fusobacterium;__	0.520 (0.494-0.547)	0.503 (0.483-0.548)	AA	MaAsLin

Abbreviations:

k\_, kingdom; p\_, phylum, c\_, class; o\_, order; f\_, family; g\_ genera; s\_, species.

AA, advanced adenoma; HC, healthy control.



**Supplementary Table S4. Fecal microbial markers associated with advanced adenoma at genera and species level shared across multiple regions in China.**

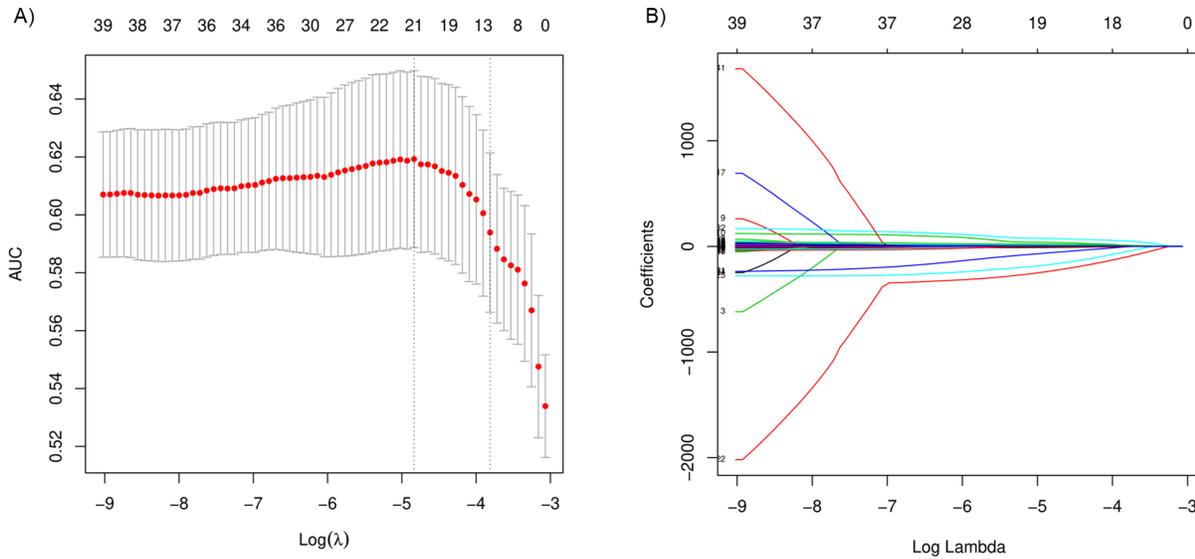
	Hunan	Anhui	Yunnan	Jiangsu	Zhejiang
k__Bacteria.p__Actinobacteria.c__Actinobacteria.o__Bifidobacteriales.f__Bifidobacteriaceae.g__Bifidobacterium.s__		√	√		
k__Bacteria.p__Actinobacteria.c__Actinobacteria.o__Micrococcales.f__Micrococcaceae.g__Rothia		√	√		
k__Bacteria.p__Firmicutes.c__Clostridia.o__Clostridiales.f__Lachnospiraceae.g__Ruminococcus_gnavusgroup.s__unculturedbacterium	√				√
k__Bacteria.p__Firmicutes.c__Clostridia.o__Clostridiales.f__Lachnospiraceae.g__Tyzzerella4	√	√			√
k__Bacteria.p__Firmicutes.c__Clostridia.o__Clostridiales.f__Lachnospiraceae.g__Tyzzerella4.s__	√	√			√
k__Bacteria.p__Firmicutes.c__Clostridia.o__Clostridiales.f__Peptostreptococcaceae.g__Terrisporobacter		√	√		
k__Bacteria.p__Firmicutes.c__Clostridia.o__Clostridiales.f__Peptostreptococcaceae.g__Terrisporobacter.s__unculturedbacterium		√	√		
k__Bacteria.p__Firmicutes.c__Clostridia.o__Clostridiales.f__Lachnospiraceae.g__Lachnospira.s		√	√		

Abbreviations: k\_, kingdom; p\_, phylum, c\_, class; o\_, order; f\_, family; g\_ genera; s\_, species.

## Supplementary figure legend

### Supplementary Fig. S1. Feature selection using the least absolute shrinkage and selection operator (LASSO) binary logistic regression model.

- A) Tuning parameter ( $\lambda$ ) selection in the LASSO model used 10-fold cross-validation via minimum criteria. The area under the receiver operating characteristic (AUC) curve was plotted versus  $\log(\lambda)$ . Dotted vertical lines were drawn at the optimal values by using the minimum criteria and the 1 standard error of the minimum criteria (the 1-SE criteria).
- B) LASSO coefficient profiles of the 49 features. A coefficient profile plot was produced against the  $\log(\lambda)$  sequence.



**Supplementary Fig. S1. Feature selection using the least absolute shrinkage and selection operator (LASSO) binary logistic regression model.**

(A) Tuning parameter ( $\lambda$ ) selection in the LASSO model used 10-fold cross-validation via minimum criteria. The area under the receiver operating characteristic (AUC) curve was plotted versus  $\log(\lambda)$ . Dotted vertical lines were drawn at the optimal values by using the minimum criteria and the 1 standard error of the minimum criteria (the 1-SE criteria).

(B) LASSO coefficient profiles of the 49 features. A coefficient profile plot was produced against the  $\log(\lambda)$  sequence.