University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

Supplemental Methods:

Faecal 16S rRNA Gene Sequencing:

Amplicon Sequencing: The V3-V4 variable region of the 16s rRNA gene was amplified from each extracted DNA sample according to the 16S metagenomic sequencing library protocol (Illumina, Sweden). Initially, the template DNA was amplified using primers specific to the V3-V4 region of the 16s rRNA gene which also incorporates the Illumina overhang adaptor (5'-3': tcgtcggcagcgtcagatgtgtataagagacagcctacgggnggcwgcag; 5'-3': gtctcgtgggctcggagatgtgtataagagacaggactachvgggtatctaatcc. Each PCR reaction contained 2.5µl DNA template, 5µl forward primer (1µM), 5µl reverse primer (1µM) (Sigma, Ireland) and 12.5µl Kapa HiFi HotStart ReadyMix (2X) (Kapa Biosystems, London, United Kingdom). The template DNA was amplified under the following PCR conditions: 95˚C for 3 min (initialisation); followed by 25 cycles of 95˚C for 30 sec (denaturation), 55˚C for 30 sec (annealing), 72˚C for 30 sec (elongation); followed by a final elongation period of 5 minutes. A negative control reaction whereby the DNA template was replaced with PCR grade water was employed to confirm lack of contamination, and PCR products were visualised using gel electrophoresis (1X TAE buffer, 1.5% agarose gel, 100V) post PCR reaction. Successful amplicons were then cleaned using the AMPure XP purification system (Labplan, Dublin, Ireland). A second PCR reaction was then performed using the previously amplified and purified DNA as the template. Two indexing primers (Illumina Nextera XT indexing primers, Illumina) were used per sample to allow all samples to be pooled, sequenced and subsequently identified. Each reaction contained 25µl Kapa HiFi HotStart ReadyMix (2X), 5µl template DNA, 5µl index 1 primer (N7xx), 5µl index 2 primer (S5xx) and 10µl PCR grade water. PCR conditions were the same as previously described with the samples undergoing just eight cycles instead of 25. PCR products then underwent the same electrophoresis and cleaning protocols as described above. Samples were then quantified using the Qubit 2.0 Fluorometer (Invitrogen, Carlsbad, CA, USA) in conjunction with the broad range DNA quantification assay

kit (Biosciences, Dublin, Ireland). All samples were then pooled to an equimolar concentration and the pool underwent a final cleaning step. The quality of the pool was determined using the Agilent Bioanalyser prior to sequencing. The sample pool was then denatured with 0.2 M NaOH, diluted to 4pM and combined with 10% (v/v) denatured 4pM PhiX. Samples were then sequenced on the MiSeq sequencing platform using a 2.300 cycle V3 Kit following protocols outlined by Illumina.

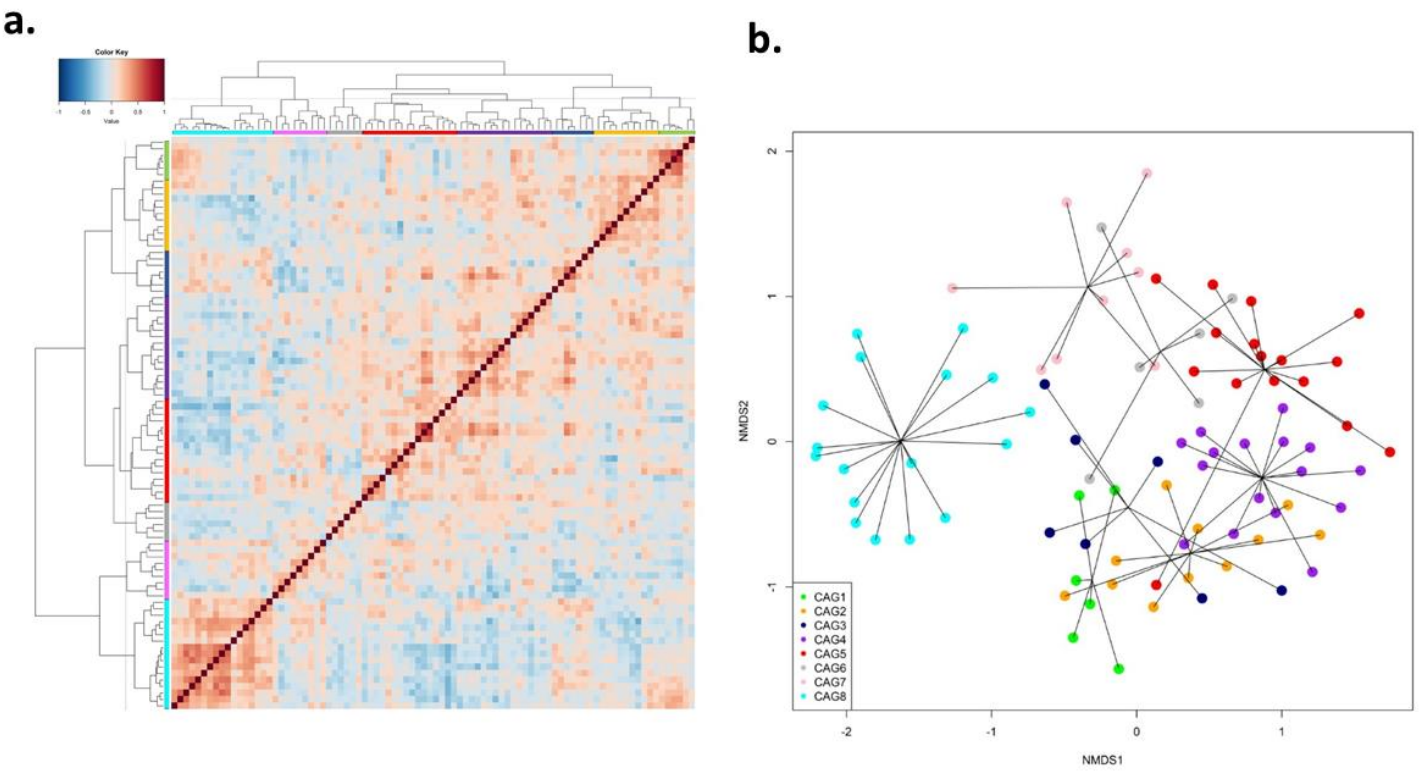
Bioinformatic and Statistical Analysis: Two-hundred and fifty base pair paired-end reads were assembled using FLASH (1). Reads were further processed with the inclusion of quality filtering, based on a quality score of >25, followed by subsequent removal of mismatched barcodes and sequences below length threshold using QIIME. USEARCH v7 (64-bit) was used for noise removal and chimera detection as well as clustering into operational taxonomic units (OTUs). At the lowest taxonomic level, reads were clustered into 100 OTUs. Sequencing generated approximately 13 million clean sequences, ranging from approximately 75000 to 395000 per sample (Supplementary file, Table 1). PyNAST was used to align OTUs, and taxonomy was assigned to OTU sequences using the Qiime2 Naïve-Bayes classifier trained on V3-V4 regions extracted from the 99% identity 16S rRNA rep set of the Silva 119 database.

Composition Analysis: The R package *compareGroups* (v. 3.1) was employed to detect statistically significant differences in abundances of individual taxa between groups using the Mann-Whitney U-test (MWU-test) with multiple corrections. Statistical significance was accepted as $p \leq 0.05$ after false discovery rate (FDR) multiple correction. Compositional alpha diversity was calculated in QIIME (v. 1.9.1) and statistical differences between groups were detected using the MWU-test. The remaining statistical analyses were all performed in R (v. 3.2.3). The *phyloseq* package (v. 1.10) was used to calculate compositional β-diversity using genus-level relative abundance data. This was visualised by principal coordinate analysis (PCoA) using *ggplot2* (v. 2.1.0). Permutational

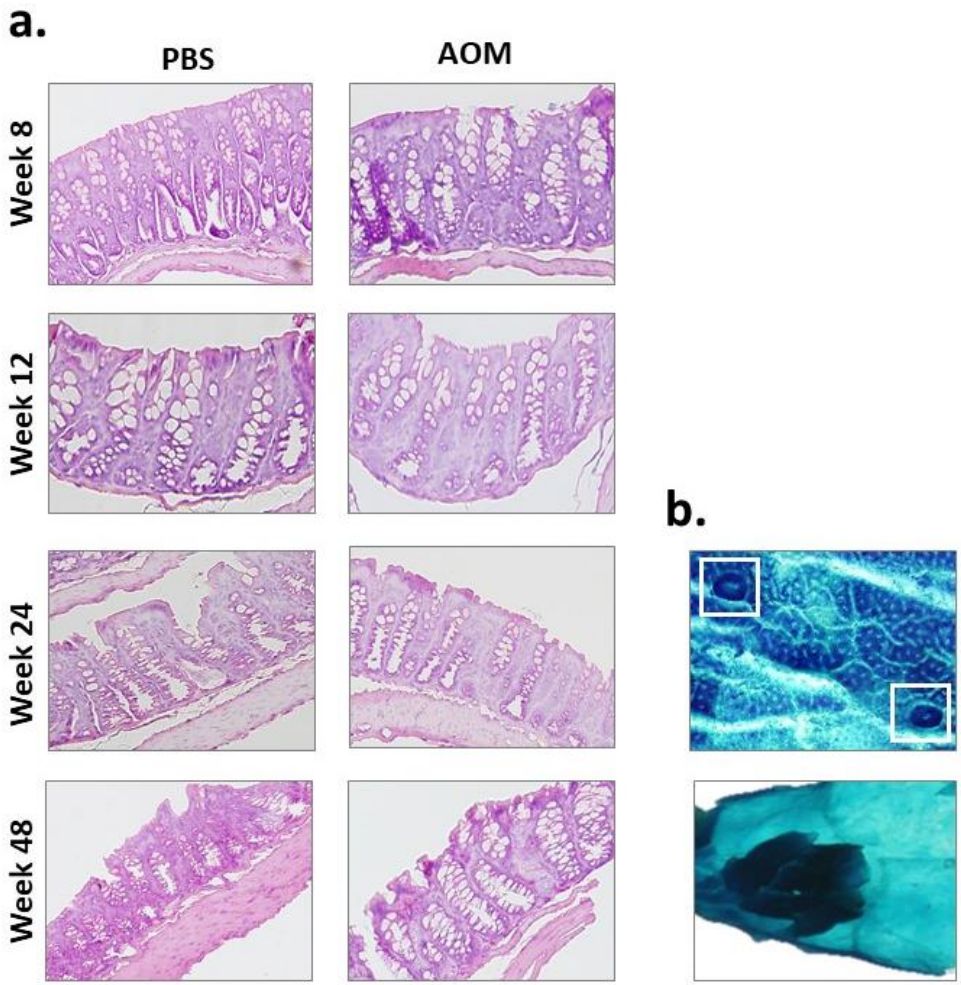
multivariate analysis of variance (PERMANOVA) was used to test for differences in overall microbiome composition between groups using the *vegan* package's 'adonis' function.

Supplemental Figures:

Supplemental Figure 1: Generation of Co-abundance Groups (CAGs). A heatplot of Kendall correlation values for each pair of taxa clustered by Ward linkage of their Pearson correlations. Coloured bars along the axes denote the CAGs (a). Non-metric multidimensional scaling (NMDS) of the distance metrics of the CAGs correlation values (b). The composition if each CAG is presented in Supplemental Table 3.



Supplemental Figure 2: Representative H&E images from PBS and AOM treated mice at each timepoint (a; 20x magnification). Methylene blue stained tissue from weeks 24 and 48 showing ACF (white boxes) and adenomas, respectively (b).



Supplemental Tables:

Supplemental Table 1: Information on sequencing result

Sample	Post QC Paired Reads
A1_PBS_48	295,122
A2_PBS_48	227,238
A3_PBS_48	248,175
A4_PBS_48	207,957
A5_PBS_48	226,384
A6_PBS_48	326,112
B1_PBS_24	252,312
B2_PBS_24	273,995
B3_PBS_24	131,959
B4_PBS_24	271,205
B5_PBS_24	313,322
B6_PBS_24	308,514
C1_PBS_12	198,378
C2_PBS_24	250,337
C3_PBS_12	259,375
C5_PBS_48	155,774
C6_PBS_48	217,138
D1_PBS_12	224,150
D2_PBS_12	257,442
D3_PBS_12	205,023
D4_PBS_12	75,233
D5_PBS_12	107,419
D6_PBS_12	255,268
E1_PBS_8	237,996

E2_PBS_8	178,569
E3_PBS_8	171,261
E4_PBS_8	142,438
E5_PBS_8	149,116
E6_PBS_8	157,760
F1_PBS_8	194,546
F2_PBS_8	191,113
G1_AOM_48	189,926
G2_AOM_48	82,468
G3_AOM_48	158,337
G4_AOM_48	173,334
G5_AOM_48	168,196
G6_AOM_48	274,927
H1_AOM_24	225,750
H2_AOM_24	341,234
H3_AOM_24	216,984
H4_AOM_24	243,222
H5_AOM_24	358,867
H6_AOM_24	277,854
i1_AOM_12	258,762
i2_AOM_24	108,535
i3_AOM_12	244,568
i4_AOM_24	281,743
i5_AOM_48	336,154
i6_AOM_48	168,177
J1_AOM_12	174,459
J2_AOM_12	253,445
J3_AOM_12	194,532
J4_AOM_12	242,148

J5_AOM_12	206,206
K1_AOM_8	212,231
K2_AOM_8	219,256
K3_AOM_8	103,886
K4_AOM_8	184,036
K5_AOM_8	213,128
K6_AOM_8	239,244
L1_AOM_8	353,915
L2_AOM_8	395,619

Supplemental Table 2: Primer Sequences and matching Universal Probe Library (UPL) Number

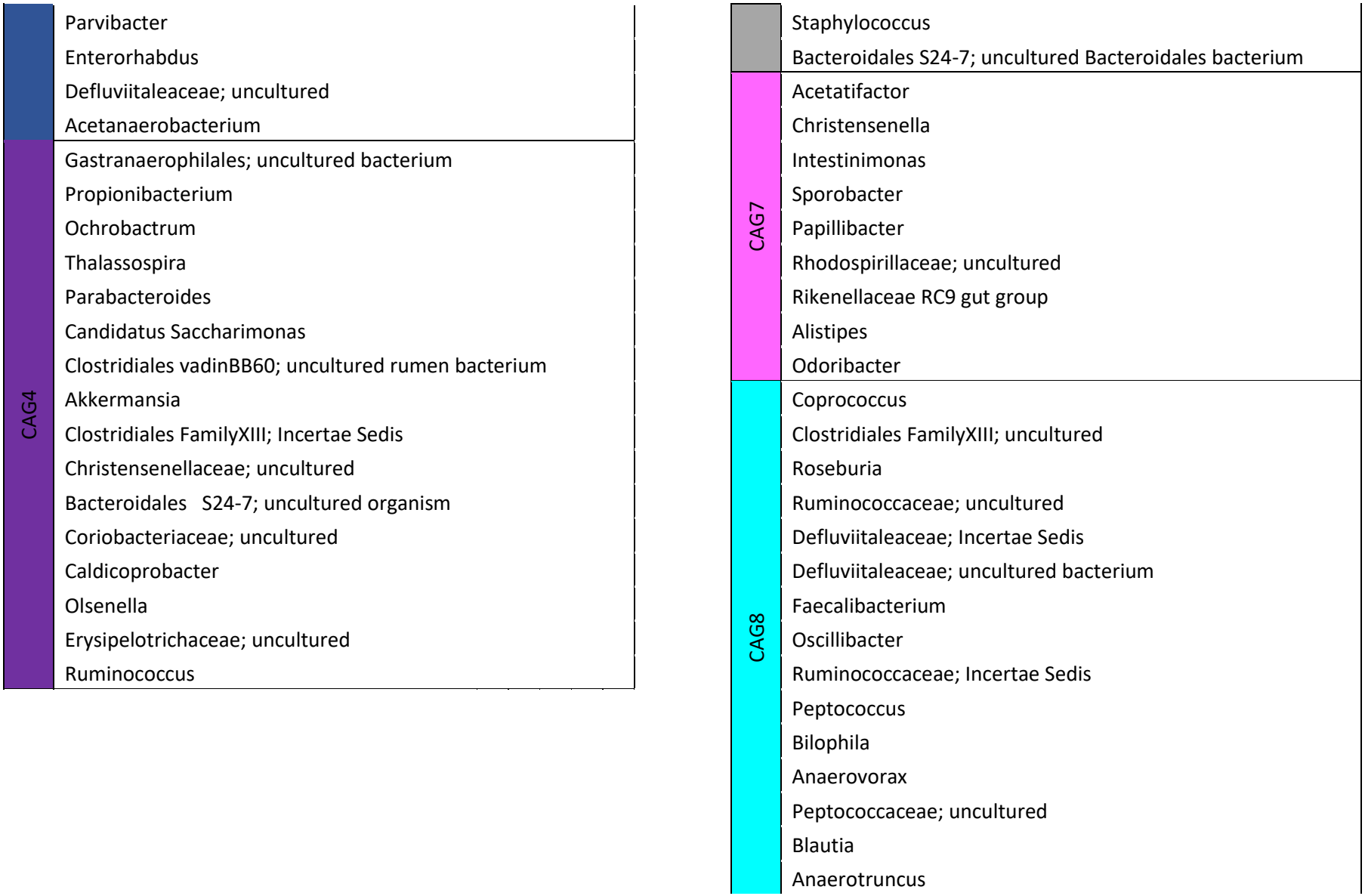
Gene	**Forward Primer (5'-3')**	**Reverse Primer (5'-3')**	**UPL Number**
Mouse			
ASBT	agctggtcaaccctggtaca	gggggagaaggagagctg	99
TNFα	ctgtagcccacgtcgtagc	ttgagatccatgccgttg	25
IL-1β	agttgacggaccccaaaag	agctggatgctctcatcagg	38
IL-6	gctaccaaactggatataatcagga	ccaggtagctatggtactccagaa	6
IL-12	aaggaacagtgggtgtccag	gttagcttctgaggacacatcttg	27
IL-10	cagagccacatgctcctaga	tgtccagctggtcctttgtt	41
CXCL1	gactccagccacactccaac	tgacagcgcagctcattg	83
CXCL2	aaaatcatccaaaagatactgaacaa	ctttggttcttccgttgagg	26
CXCL5	tagagccccaatctccacac	gagctggaggctcattgtg	67
TGF-β	tggagcaacatgtggaactc	gtcagcagccggttacca	72
β-actin	ctaaggccaaccgtgaaaag	accagaggcatacagggaca	64
Human			
TNFα	cgctccccaagaagacag	agaggctgaggaacaagcac	57
IL-6	caggagcccagctatgaact	agcaggcaacaccaggag	7
IL-8	agacagcagagcacacaagc	atggttccttccggtggt	72
IL-10	cataaattagaggtctccaaaatcg	cataaattagaggtctccaaaatcg	45
CXCL2	cccatggttaagaaaatcatcg	cttcaggaacagccaccaat	69

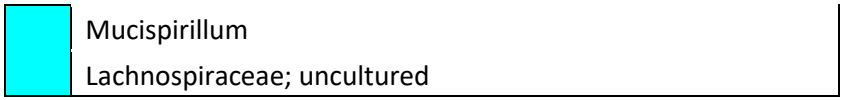
CXCL5	cagcgctctcttgaccacta	cacaaggagctcgaaggacc	28
TGF-β	actactacgccaaggaggtcac	tgcttgaacttgtcatagatttcg	31
β-actin	ccagaggcgtacagggat	ccaaccgcgagaagatga	64

Supplemental Table 3: Composition of CAGS

CAG1	Citrobacter
	Hydrogenoanaerobacterium
	Clostridiales vadinBB60; unidentified
	Clostridiales vadinBB60; uncultured Clostridia bacterium
	Clostridiales vadinBB60; uncultured bacterium
	Anaeroplasma
CAG2	Marvinbryantia
	Rikenella
	Anaerofustis
	Lactobacillus
	Bacteroides
	Enterococcus
	Porphyromonadaceae; uncultured
	Psychrobacter
	Vibrio
	Bacteroidales S24-7; mouse gut metagenome
	Prevotellaceae; uncultured
CAG3	Escherichia-Shigella
	Lachnospiraceae; Incertae Sedis
	Desulfovibrio

CAG5	Streptococcus
	Bacteroidales S24-7; uncultured bacterium
	Mollicutes RF9; uncultured Erysipelotrichaceae bacterium
	Mollicutes RF9; uncultured Firmicutes bacterium
	Allobaculum
	Bifidobacterium
	Parasutterella
	Firmicutes bacterium CAG822
	Turicibacter
	Mollicutes RF9; uncultured rumen bacterium
	Clostridium sensu stricto 1
	Gordonibacter
	Mollicutes RF9; uncultured Mollicutes bacterium
	Mollicutes RF9; uncultured bacterium
	Mollicutes RF9; unidentified
	Mollicutes RF9; uncultured Paenibacillaceae bacterium
CAG6	Lachnospira
	Anaerobacillus
	Delftia
	Candidatus Arthromitus


Parvibacter
Enterorhabdus
Defluviitaleaceae; uncultured
Acetanaerobacterium
CAG4
Gastranaerophilales; uncultured bacterium
Propionibacterium
Ochrobactrum
Thalassospira
Parabacteroides
Candidatus Saccharimonas
Clostridiales vadinBB60; uncultured rumen bacterium
Akkermansia
Clostridiales FamilyXIII; Incertae Sedis
Christensenellaceae; uncultured
Bacteroidales S24-7; uncultured organism
Coriobacteriaceae; uncultured
Caldicoprobacter
Olsenella
Erysipelotrichaceae; uncultured
Ruminococcus
Staphylococcus
Bacteroidales S24-7; uncultured Bacteroidales bacterium
CAG7
Acetatifactor
Christensenella
Intestinimonas
Sporobacter
Papillibacter
Rhodospirillaceae; uncultured
Rikenellaceae RC9 gut group
Alistipes
Odoribacter
CAG8
Coprococcus
Clostridiales FamilyXIII; uncultured
Roseburia
Ruminococcaceae; uncultured
Defluviitaleaceae; Incertae Sedis
Defluviitaleaceae; uncultured bacterium
Faecalibacterium
Oscillibacter
Ruminococcaceae; Incertae Sedis
Peptococcus
Bilophila
Anaerovorax
Peptococcaceae; uncultured
Blautia
Anaerotruncus


Mucispirillum
Lachnospiraceae; uncultured

Supplemental Table 4: Spearman correlation analyses between individual taxa and bile acids at weeks 12 and 24.

Week 12			
P-value	**Taxon**	**Bile acid**	**R-value**
<0.001	Tenericutes	T-UDCA	-1
<0.001	Verrucomicrobia	UDCA	-1
<0.001	*Parasutterella*	T-UDCA	1
<0.001	*Escherichia_Shigella*	DCA	-1
<0.001	*Akkermansia*	UDCA	-1
Week 24			
<0.001	Tenericutes	T-DCA	-1
<0.001	Verrucomicrobia	T-CA	-1
<0.001	Verrucomicrobia	T-UDCA	-1
<0.001	*Coriobacteriaceae_uncultured*	Total BA	-1
<0.001	*Bacteroides*	DCA	-1
<0.001	*Caldicoprobacter*	T-CA	-1
<0.001	*Caldicoprobacter*	T-UDCA	-1
<0.001	*Christensenellaceae_uncultured*	T-CDCA	-1
<0.001	*Akkermansia*	T-CA	-1
<0.001	*Akkermansia*	T-UDCA	-1

References

1. Magoč T, Salzberg SL. FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics. 2011 Nov 1;27(21):2957-63.