Locoregional effects of microbiota in a preclinical model of colon carcinogenesis

Sarah Tomkovich[1,2], Ye Yang[1], Kathryn Winglee[5], Josee Gauthier[1], Marcus Mühlbauer[1], Xiaolun Sun[1], Mansour Mohamadzadeh[§], Xiuli Liu[6], Patricia Martin[3,4], Gary P. Wang[7], Eric Oswald [3,4], Anthony A. Fodor[5] and Christian Jobin[1§*]

[1]Department of Medicine, [§]Department of Infectious Diseases and Pathology, University of Florida, Gainesville, FL 32611, USA. [2]Department of Microbiology and Immunology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27713, USA. [3]IRSD, Université de Toulouse, INSERM, INRA, ENVT, UPS, Toulouse, France. [4]CHU Toulouse, Service de Bactériologie-Hygiène, Toulouse, France. [5]Department of Bioinformatics and Genomics, University of North Carolina at Charlotte, North Carolina 28223, United States. [6]Department of Pathology, Immunology, and Laboratory Medicine, University of Florida, Gainesville, FL 32611, USA. [7]Department of Medicine, Division of Infectious Diseases and Global Medicine, University of Florida, Gainesville, FL 32611, USA. [*]Corresponding author

Supplementary Materials and Methods

16S rRNA sequencing analysis

The forward and reverse reads were stitched together using fastq-join (1). Taxonomic ranks were assigned for the stitched reads using the RDP (ribosomal database project) classifier (2) version 2.2 with confidence set to 80%. Reads were grouped by genera and the counts were normalized and $\log_{10}$ transformed (3) using the following formula:

$$log_{10}\left(\frac{RC}{n} \text{ x } \frac{\sum x}{N} + 1\right)$$

where RC is the read count for a particular OTU in a particular sample, *n* is the total number of reads in that sample, the sum of *x* is the total number of reads in all samples and N is the total number of samples. The Principle Coordinate Analysis (PCoA) was generated from the Bray-Curtis distance of the normalized and $\log_{10}$ transformed counts using the capscale function in the vegan R package (4,5).

Genera significantly associated with colon tumor number or combined inflammation score were detected using the gls function in the R nlme package, with the REML method (6) to fit a generalized mixed linear model of the form:

$$\text{genera} \sim \text{variable} + 1|\text{cage} + \varepsilon$$

where genera indicates the $\log_{10}$ normalized abundance of a particular genera, variable indicates either the number of colon tumors or the combined inflammation score and 1|cage indicates that we used the cage as a random effect. We filtered genera absent in more than a quarter of the samples. We then ran an ANOVA analysis on the above model to generate p-values for each variable. The p-values for cage were calculated using an ANOVA to compare this model and a model with the cage removed (genera ~

variable + ε). The p-values were then adjusted for multiple hypothesis testing using the method of Benjamini & Hochberg (7). The Spearman correlation was calculated using the R function cor (5) with x as the number of colon tumors or the combined inflammation score and y as $\log_{10}$ normalized genera abundance. The code and tables used to generate the 16S rRNA sequencing figure and supplementary table can be found at: https://github.com/afodor/apcmin.

References

1. Aronesty E. ea-utils : “Command-linetools for processing biological sequencing data” [Internet]. 2011 [cited 2016 Nov 4]. Available from: https://github.com/ExpressionAnalysis/ea-utils

2. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol. 2007;73:5261–7.

3. McCafferty J, Mühlbauer M, Gharaibeh RZ, Arthur JC, Perez-Chanona E, Sha W, et al. Stochastic changes over time and not founder effects drive cage effects in microbial community assembly in a mouse model. ISME J. 2013;7:2116–25.

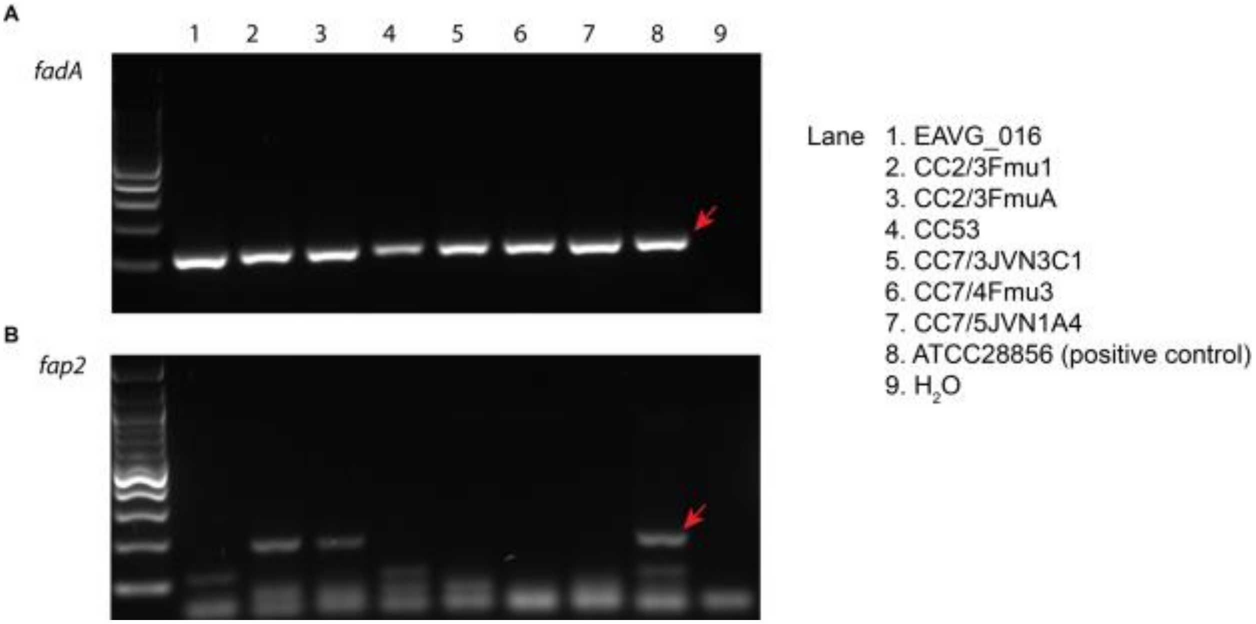
4. Oksanen JB, Guillaume F, Kindt R, Legendre P, Minchin PR, O’Hara RB, et al. vegan: Community Ecology Package. 2015.

5. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing; 2015.

6. Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core Team. nlme: Linear and Nonlinear Mixed Effects Models. 2016.

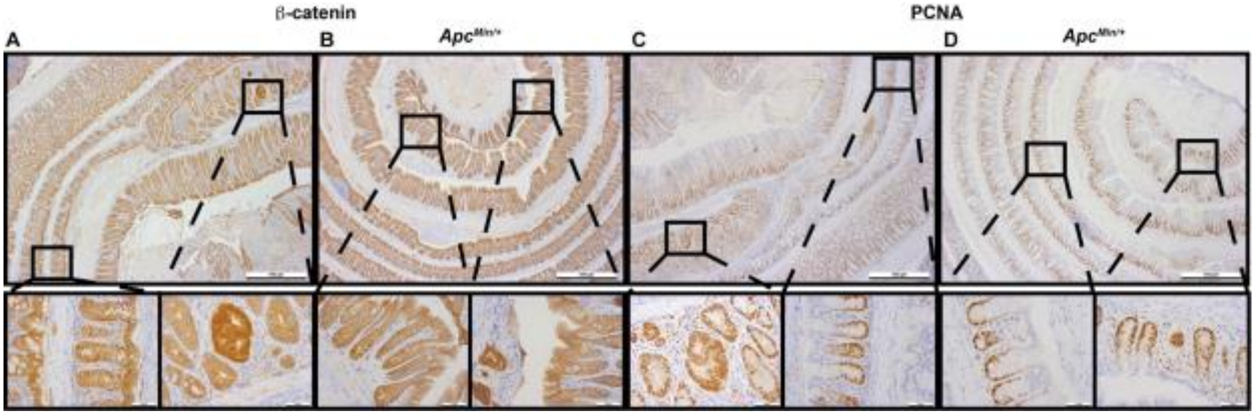
7. Benjamini Y, Hochberg Y. Controlling the false discovery rate: apractical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B (Methodological). 1995;57:289–300.

Supplementary Figure S1.



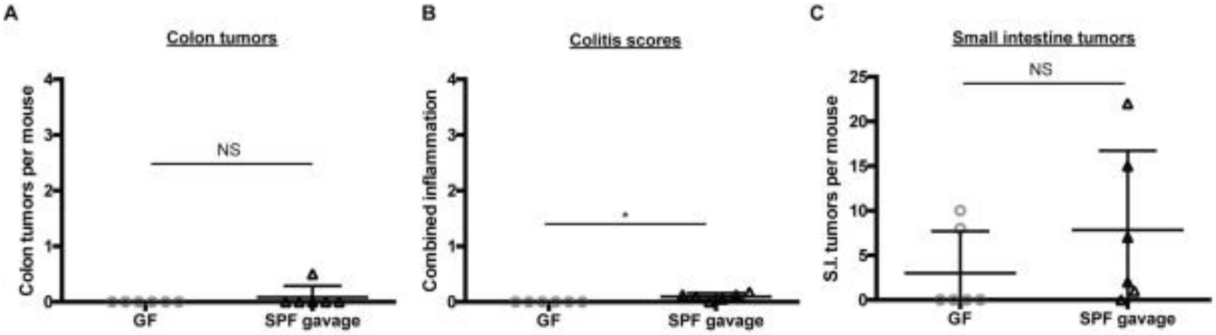
Supplementary Figure S1. PCR detection of *F. nucleatum* adhesins. **A-B**) Detection of the *fadA* (**A**) and *fap2* (**B**) adhesins in tested *F. nucleatum* strains.

Supplementary Figure S2.



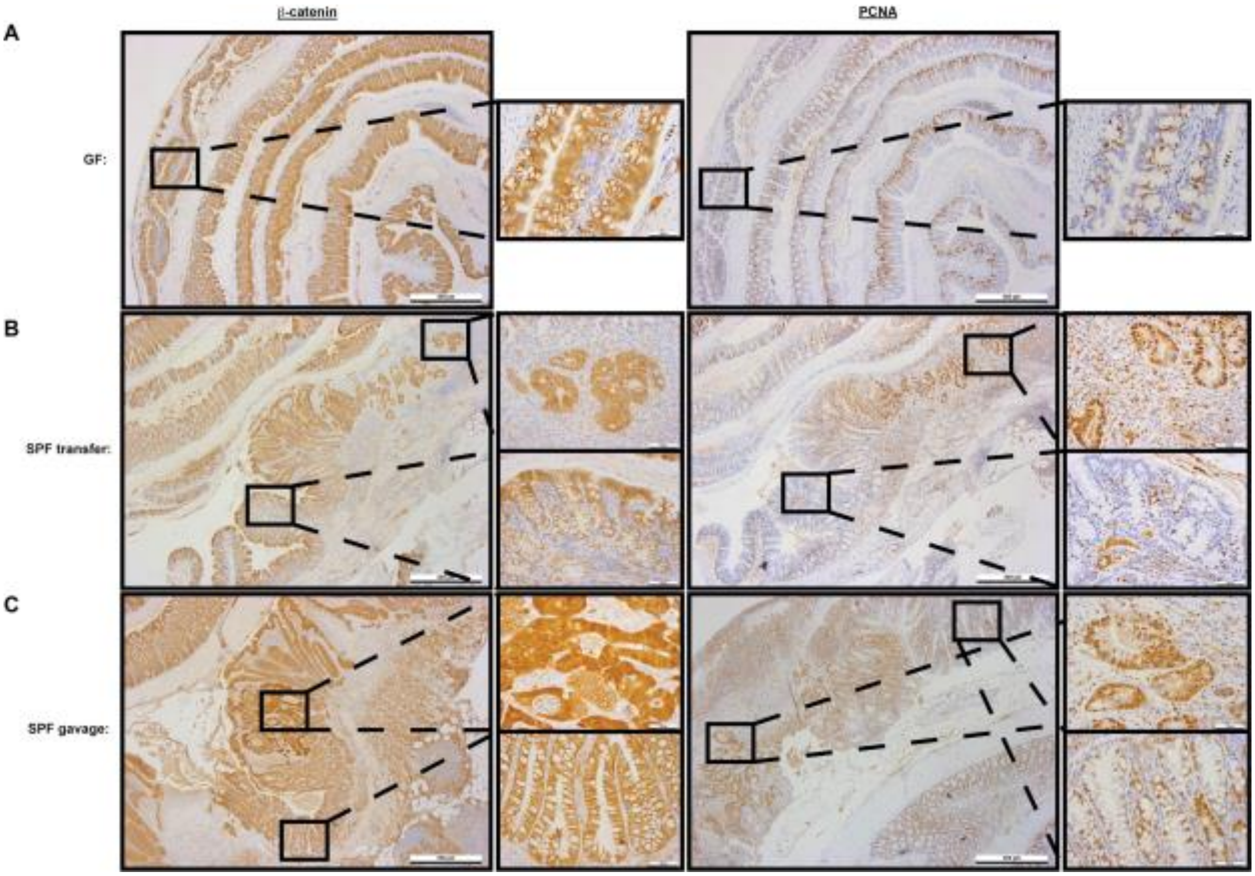
Supplementary Figure S2. *Apc*$^{Min/+}$;*Il10*$^{-/-}$ mice have increased colon proliferation. **A-B**) CTNNB1 immunohistochemistry (IHC) from 30-40 week old SPF *Apc*$^{Min/+}$;*Il10*$^{-/-}$ (**A**) and *Apc*$^{Min/+}$ (**B**) colons. **C-D**) PCNA IHC from SPF *Apc*$^{Min/+}$;*Il10*$^{-/-}$ (**C**) and *Apc*$^{Min/+}$ (**D**) colons. Higher magnification of both dysplastic and normal regions are shown for SPF *Apc*$^{Min/+}$;*Il10*$^{-/-}$ mice.

Supplementary Figure S3.



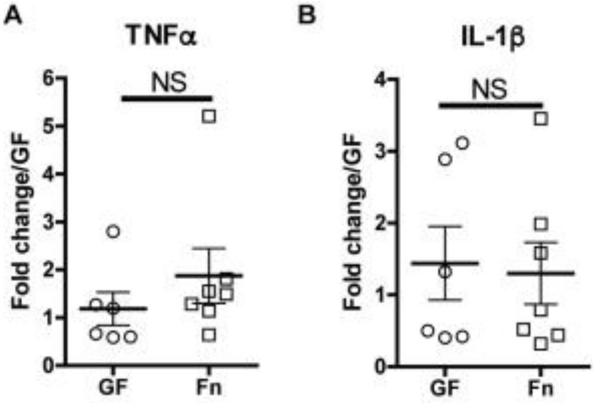
Supplementary Figure S3. Bacteria do not promote colon inflammation and tumorigenesis in *Apc*$^{Min/+}$ mice. **A-C**) Colon macroscopic tumor counts (**A**), colitis scores (**B**), and small intestine tumor counts (**C**) from GF (n=6) and SPF gavaged (n=6) *Apc*$^{Min/+}$mice. Data are expressed as mean +/- standard deviation with *p< 0.05, or NS: not significant (p> 0.05) shown.

Supplementary Figure S4.



Supplementary Figure S4. Microbiota promote proximal colon proliferation in $Apc^{Min/+}$;$Il10^{-/-}$ mice. **A-C**) β-catenin and PCNA IHC from GF (**A**), SPF transferred (**B**), and SPF gavaged (**C**) $Apc^{Min/+}$;$Il10^{-/-}$ proximal colons. Higher magnification of both dysplastic and normal regions are shown for SPF transfer and gavage $Apc^{Min/+}$;$Il10^{-/-}$ mice.

Supplementary Figure S5.



Supplementary Figure S5. Colibactin promotes CRC development in AOM/*Il10*$^{-/-}$ mice. Colon tumor counts in 20 week *E. coli* NC101 (n=6) or Δ*clbP* (n=5) mono-associated AOM/*Il10*$^{-/-}$ mice. Data are expressed as mean +/- standard deviation with $^*p < 0.05$.

Supplementary Table S2. V1-V3 MiSeq Primers

Name	Sequence
PE1_27F-1	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT**AGTA**AGAGTTTGATCCTGGCTCAG
PE1_27F-2	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT**TCAT**AGAGTTTGATCCTGGCTCAG
PE1_27F-3	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT**GTCT**AGAGTTTGATCCTGGCTCAG
PE1_27F-4	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT**TCAAG**AGAGTTTGATCCTGGCTCAG
PE1_27F-5	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT**CTGAT**AGAGTTTGATCCTGGCTCAG
PE2_534R-1	CAAGCAGAAGACGGCATACGAGATCGGCATTCCTGCTGAACCGCTCTTCCGATCT**AAGC**ATTACCGCGGCTGCTGG
PE2_534R-2	CAAGCAGAAGACGGCATACGAGATCGGCATTCCTGCTGAACCGCTCTTCCGATCT**TACC**ATTACCGCGGCTGCTGG
PE2_534R-3	CAAGCAGAAGACGGCATACGAGATCGGCATTCCTGCTGAACCGCTCTTCCGATCT**CTAC**ATTACCGCGGCTGCTGG
PE2_534R-4	CAAGCAGAAGACGGCATACGAGATCGGCATTCCTGCTGAACCGCTCTTCCGATCT**GCAGT**ATTACCGCGGCTGCTGG
PE2_534R-5	CAAGCAGAAGACGGCATACGAGATCGGCATTCCTGCTGAACCGCTCTTCCGATCT**TAGCT**ATTACCGCGGCTGCTGG
PE2_534R-6	CAAGCAGAAGACGGCATACGAGATCGGCATTCCTGCTGAACCGCTCTTCCGATCT**AGTGA**ATTACCGCGGCTGCTGG
PE2_534R-7	CAAGCAGAAGACGGCATACGAGATCGGCATTCCTGCTGAACCGCTCTTCCGATCT**GTAGTG**ATTACCGCGGCTGCTGG
PE2_534R-8	CAAGCAGAAGACGGCATACGAGATCGGCATTCCTGCTGAACCGCTCTTCCGATCT**CATGCG**ATTACCGCGGCTGCTGG
PE2_534R-9	CAAGCAGAAGACGGCATACGAGATCGGCATTCCTGCTGAACCGCTCTTCCGATCT**GACTGT**ATTACCGCGGCTGCTGG
PE2_534R-10	CAAGCAGAAGACGGCATACGAGATCGGCATTCCTGCTGAACCGCTCTTCCGATCT**CACTGT**ATTACCGCGGCTGCTGG
PE2_534R-11	CAAGCAGAAGACGGCATACGAGATCGGCATTCCTGCTGAACCGCTCTTCCGATCT**CAGAGCT**ATTACCGCGGCTGCTGG
PE2_534R-12	CAAGCAGAAGACGGCATACGAGATCGGCATTCCTGCTGAACCGCTCTTCCGATCT**AGCATGT**ATTACCGCGGCTGCTGG
PE2_534R-13	CAAGCAGAAGACGGCATACGAGATCGGCATTCCTGCTGAACCGCTCTTCCGATCT**TATCGTG**ATTACCGCGGCTGCTGG
PE2_534R-14	CAAGCAGAAGACGGCATACGAGATCGGCATTCCTGCTGAACCGCTCTTCCGATCT**GTACATC**ATTACCGCGGCTGCTGG
PE2_534R-15	CAAGCAGAAGACGGCATACGAGATCGGCATTCCTGCTGAACCGCTCTTCCGATCT**GTCAGCAT**ATTACCGCGGCTGCTGG
PE2_534R-16	CAAGCAGAAGACGGCATACGAGATCGGCATTCCTGCTGAACCGCTCTTCCGATCT**TAGTCACG**ATTACCGCGGCTGCTGG
PE2_534R-17	CAAGCAGAAGACGGCATACGAGATCGGCATTCCTGCTGAACCGCTCTTCCGATCT**ACGAGTGC**ATTACCGCGGCTGCTGG
PE2_534R-18	CAAGCAGAAGACGGCATACGAGATCGGCATTCCTGCTGAACCGCTCTTCCGATCT**GACCACTT**ATTACCGCGGCTGCTGG

Unique barcode in red.

Supplementary Table S2. Spearman correlations and corresponding p values from linear models used to generate Figure 3.

Genus	Colon Tumor Number Spearman Correlation	P-value for Colon Tumor Number	P-value for Cage in Mixed Model with Colon Tumor Number	Adjusted P-value for Colon Tumor Number	Adjusted P-value for Cage in Mixed Model with Colon Tumor Number	Combined Inflammation Score Spearman Correlation	P-value for Combined Inflammation Score	P-value for Cage in Mixed Model with Inflammation Score	Adjusted P-value for Combined Inflammation Score	Adjusted P-value for Cage in Mixed Model with Inflammation Score
Akkermansia	0.35	0.0175	7.37E-08	0.0379	2.87E-07	0.36	0.1848	4.94E-07	0.2325	1.48E-06
Alistipes	-0.03	0.0444	1.38E-13	0.0693	1.80E-12	0.02	0.0094	4.90E-14	0.0194	6.36E-13
Allobaculum	0.35	2.08E-08	0.0013	8.10E-07	0.0019	0.27	0.0019	0.0027	0.0049	0.0036
Anaeroplasma	-0.21	0.0938	1.38E-07	0.1307	4.90E-07	-0.26	0.1074	2.13E-07	0.1551	7.54E-07
Anaerostipes	-0.40	0.0192	0.0028	0.0395	0.0039	-0.46	0.0002	0.0084	0.0009	0.0109
Anaerotruncus	-0.42	0.0014	0.1940	0.0062	0.2162	-0.32	0.0003	0.0409	0.0013	0.0489
Anaerovorax	0.18	0.8389	2.10E-05	0.8389	4.31E-05	-0.04	0.0500	5.02E-06	0.0813	1.15E-05
Bacteroides	0.36	0.8199	0	0.8389	0	0.29	0.6818	0	0.7386	0
Barnesiella	0.03	0.6268	0.0001	0.6607	0.0003	0.04	0.9242	0.0002	0.9485	0.0003
Bifidobacterium	0.18	0.1295	0.0008	0.1742	0.0012	0.12	0.3854	0.0014	0.4294	0.0020
Blautia	0.15	0.0044	6.84E-08	0.0123	2.87E-07	0.28	5.07E-05	8.45E-09	0.0003	3.71E-08
Butyricicoccus	-0.27	0.0171	3.78E-06	0.0379	9.22E-06	-0.32	0.0052	1.26E-06	0.0112	3.27E-06
Butyrivibrio	-0.65	5.75E-05	4.03E-05	0.0004	7.48E-05	-0.52	0.0002	1.86E-05	0.0009	3.81E-05
Clostridium IV	-0.43	5.68E-05	2.10E-06	0.0004	6.26E-06	-0.40	1.47E-05	3.91E-07	0.0002	1.27E-06
Clostridium XI	0.52	0.0012	0.0001	0.0056	0.0003	0.46	0.0008	0.0002	0.0026	0.0003
Clostridium XlVa	-0.29	0.0228	0.1387	0.0434	0.1591	-0.26	0.0016	0.1370	0.0045	0.1522
Clostridium XlVb	-0.45	0.0245	1.17E-06	0.0434	3.81E-06	-0.39	1.59E-05	8.55E-09	0.0002	3.71E-08
Clostridium XVIII	-0.01	0.5614	4.28E-10	0.6081	2.78E-09	0.17	0.1435	5.81E-10	0.1930	3.24E-09

Coprobacillus	0.19	0.0237	4.39E-08	0.0434	2.14E-07	0.17	0.1271	1.16E-07	0.1771	4.52E-07
Coprococcus	-0.19	0.1928	0.0466	0.2425	0.0551	-0.22	0.0490	0.0414	0.0813	0.0489
Dorea	0.29	0.0138	0.0264	0.0336	0.0332	0.30	0.0161	0.0519	0.0298	0.0595
Enterococcus	0.61	8.03E-06	0.0002	7.83E-05	0.0004	0.40	0.0003	0.0002	0.0013	0.0003
Enterorhabdus	-0.50	0.0419	2.52E-06	0.0693	6.55E-06	-0.43	0.0049	6.31E-07	0.0112	1.76E-06
Erysipelotrichaceae incertae sedis	-0.20	0.5089	0.0034	0.5670	0.0046	-0.10	0.3202	0.0012	0.3785	0.0017
Escherichia/Shigella	0.35	0.0016	3.90E-05	0.0062	7.48E-05	0.18	0.0999	2.72E-05	0.1498	5.31E-05
Flavonifractor	-0.19	0.1593	1.54E-11	0.2071	1.50E-10	-0.14	0.0004	3.48E-13	0.0013	3.39E-12
Lachnoanaerobaculum	-0.46	0.0017	0.9284	0.0062	0.9284	-0.33	0.0050	0.9352	0.0112	0.9352
Lachnospiracea incertae sedis	-0.51	0.0022	1.57E-09	0.0072	8.77E-09	-0.39	0.0012	3.06E-10	0.0035	1.99E-09
Lactobacillus	0.12	0.4745	7.86E-11	0.5443	6.13E-10	0.18	0.9804	1.26E-10	0.9804	9.79E-10
Marvinbryantia	-0.51	0.0044	0.0004	0.0123	0.0006	-0.24	0.1580	0.0001	0.2054	0.0002
Oscillibacter	-0.70	3.37E-06	0.0454	4.39E-05	0.0551	-0.63	1.04E-09	0.0196	4.07E-08	0.0246
Pseudoflavonifractor	-0.61	1.03E-06	0.5268	2.02E-05	0.5553	-0.58	5.38E-09	0.1405	1.05E-07	0.1522
Robinsoniella	0.29	0.0073	3.22E-15	0.0189	6.28E-14	0.15	0.0669	3.89E-15	0.1043	7.58E-14
Roseburia	-0.18	0.0441	8.28E-06	0.0693	1.90E-05	-0.25	0.0134	9.60E-06	0.0260	2.08E-05
Ruminococcus2	-0.31	0.0866	1.62E-05	0.1251	3.52E-05	-0.40	0.0310	8.24E-05	0.0550	0.0002
Staphylococcus	-0.16	0.2232	0.8027	0.2720	0.8238	-0.12	0.3724	0.6830	0.4271	0.7010
Streptococcus	-0.28	0.0745	0.3529	0.1118	0.3823	-0.23	0.2006	0.5401	0.2445	0.5693
Syntrophococcus	-0.62	0.0001	0.0039	0.0007	0.0051	-0.49	2.57E-05	0.0001	0.0002	0.0002
Turicibacter	0.21	0.2970	2.25E-06	0.3510	6.26E-06	-0.01	0.7483	2.34E-06	0.7887	5.70E-06

Supplementary Table S3. Accession numbers for 16S rRNA sequenced *$Apc^{Min/+}$*;*$Il10^{-/-}$* stool samples from Figure 3.

Sample ID	SRA accession number	number of reads	cage	colon tumor number	combined inflammation score
1	SRR4453219	288706	54802	1	0.5
2	SRR4453218	273580	637823	2	0.875
3	SRR4453243	239102	51541	8	2.875
5	SRR4453220	388330	52048	2	0
6	SRR4453246	288676	54253	8	2.5
7	SRR4453208	419434	56547	15	3.125
8	SRR4453217	300654	584724	2	0.75
10	SRR4453213	192538	53690	1	0.875
11	SRR4453248	247728	55908	3	1.375
12	SRR4453223	241112	52561	3	2.125
13	SRR4453252	318654	55906	1	0.375
14	SRR4453251	257246	644039	22	3.25
15	SRR4453211	257872	55907	0	0.75
17	SRR4453185	308250	578031	7	2.25
18	SRR4453245	234064	55908	3	0.875
19	SRR4453244	250950	55907	0	0.5
20	SRR4453214	279160	51541	1	0.25
21	SRR4453247	317838	56800	2	0
22	SRR4453206	224598	56800	2	1.75
23	SRR4453212	232396	50925	2	0.5
25	SRR4453210	266526	51542	0	0
26	SRR4453216	182184	50927	2	0.5
27	SRR4453241	252032	54252	2	0.5
28	SRR4453242	247056	54803	2	0
29	SRR4453205	238380	605881	6	2.375
30	SRR4453222	221232	54530	5	1.375
31	SRR4453224	262342	578031	13	2.5
33	SRR4453250	264660	51542	0	0.25
34	SRR4453202	219824	578030	2	0
35	SRR4453207	139938	625434	6	2.25
36	SRR4453249	247550	50927	1	0.25
37	SRR4453196	207336	669929	21	3.25
38	SRR4453195	211686	584724	1	2
39	SRR4453194	206174	55906	0	1.75
40	SRR4453201	215064	56548	9	3.25
41	SRR4453200	194462	56800	1	0.5
42	SRR4453199	191006	50926	10	3.875
43	SRR4453198	196080	50925	2	0.625
44	SRR4453204	192844	54802	6	2.75
45	SRR4453203	194830	55907	9	3.5
46	SRR4453235	230700	56801	2	0.375
47	SRR4453236	233494	52048	10	2.5
49	SRR4453215	271316	50928	1	1.875
51	SRR4453221	184994	52933	0	0.375
52	SRR4453239	203852	52048	12	0.75
53	SRR4453240	213666	55906	1	0.25
54	SRR4453237	200366	637823	2	0.375
55	SRR4453238	218774	54252	1	0.25

56	SRR4453209	190800	56548	7	3.25
57	SRR4453197	301744	578030	0	0
58	SRR4453187	205868	639275	6	2.25
59	SRR4453186	240942	54530	7	2.5
61	SRR4453189	207414	53690	0	0.75
62	SRR4453188	223306	584725	2	1.1
63	SRR4453191	250208	52560	5	1.75
65	SRR4453190	283568	50928	0	0.75
66	SRR4453193	238832	639275	4	1.625
67	SRR4453192	216246	605881	3	1.75
68	SRR4453184	290478	56548	7	2.25
70	SRR4453183	223168	53691	1	0
71	SRR4453225	867326	53690	2	0.5
72	SRR4453226	175708	54967	0	0
73	SRR4453227	242832	584724	0	1
74	SRR4453228	220574	54967	0	0.375
75	SRR4453229	218900	644039	15	2.625
76	SRR4453230	237422	50925	6	2.5
77	SRR4453231	243032	54253	16	1.5
78	SRR4453232	210244	51542	1	0.375
79	SRR4453233	331504	53690	2	0.75
80	SRR4453234	222050	605882	0	0.25
Average		250163.1714		4.23	1.31
Standard deviation		88400.39565		5.02	1.10