

## *Supplementary Material*

— — —

# Variation in gut microbiota composition impacts host gene expression

Allison L Richards<sup>1</sup>, Amanda L Muehlbauer<sup>3,4</sup>, Adnan Alazizi<sup>1</sup>,  
Michael B Burns<sup>3,4</sup>, Trevor J Gould<sup>3,4</sup>, Camilla Cascardo<sup>1</sup>, Roger Pique-Regi<sup>1,2</sup>,  
Ran Blekhman<sup>3,4,5</sup>, Francesca Luca<sup>1,2,5</sup>

<sup>1</sup>Center for Molecular Medicine and Genetics, and

<sup>2</sup>Department of Obstetrics and Gynecology, Wayne State University, Detroit, Michigan 48201, USA

<sup>3</sup>Department of Genetics, Cell Biology and Development, and

<sup>4</sup>Department of Ecology, Evolution and Behavior, University of Minnesota, Minneapolis, Minnesota, USA

<sup>5</sup>To whom correspondence should be addressed.

Email: fluca@wayne.edu, blekhman@umn.edu, rpique@wayne.edu

October 27, 2017

## Cell culture and treatment

Experiments were conducted using primary human colonic epithelial cells (HCoEpiC, lot: 9810) which we also term, colonocytes (ScienCell 2950). Briefly, cells were cultured on plates or flasks coated with poly-L-lysine (PLL) (ScienCell 0413) in colonic epithelial cell medium supplemented with colonic epithelial cell growth supplement and penicillin/streptomycin, according to manufacturer's protocol (ScienCell 2951) at 37°C with 5% CO<sub>2</sub>. 24 hours before treatment, cells were changed to antibiotic-free media and moved to an incubator at 37°C, 5% CO<sub>2</sub>, and 5% O<sub>2</sub> to mimic the low-oxygen environment that is present in the gut.

Fecal microbiota was purchased from OpenBiome as live microbial suspension in 12.5% glycerol. Once arrived in our lab, the extract was not thawed until the day of treatment. Fecal microbiota was collected from healthy individuals (Ind. 1: 02-028-C, Ind. 2: 0065-0016-D, Ind. 3: 0110-0006-01, Ind. 4: 0111-0014-01, Ind. 5: 0112-0002-02). Prior to treatment, the fecal microbiota was thawed at 30°C and the microbial density was assessed by spectrophotometer (OD600) (Bio-Rad SmartSpec 3000). Media was removed from the colonocytes and fresh antibiotic-free media was added, followed by inoculation with microbial extract for a final microbial ratio of 10:1 microbe:colonocyte in each well. This ratio of microbe:colonocyte was utilized based off of our previous results suggesting that this ratio is best to reconstitute a symbiotic relationship [16]. Additional wells containing only colonocytes were also cultured in the same 6-well plate to be used as controls. This experimental protocol was already described in [16].

Following 1, 2 or 4 hours, the wells were scraped on ice, pelleted and washed with cold PBS and then resuspended in lysis buffer (Dynabeads mRNA Direct Kit) and stored at -80°C until extraction of colonocyte RNA.

## *Collinsella* Spike-in Experiment

*Collinsella aerofaciens* was purchased from ATCC (cat#: 25986) and grown in Reinforced Clostridial Medium (BD Biosciences, cat#: 218081) following manufacturer's protocol, in anaerobic conditions. The

culture was then centrifuged at 6,000 xg for 15 minutes, the media was removed and the pellet was resuspended in a 12.5% glycerol solution (in normal saline buffer). We verified that we were utilizing *Collinsella aerofaciens* by extracting the DNA with the PowerSoil kit as described below. We then PCR amplified the 16S region using primer specific to *Collinsella aerofaciens* with annealing temperature of 63°C, forward: 5'-CTTTCAGCAGGGAAGAGTCAA-3', reverse: 5'-AGCCATGCACCACCTGTATGG-3' [10]. The PCR product was run on agarose gel with a 100bp ladder (NEB, cat#: N3231S). We observed the expected band of 590bp (Figure S8 in the *Collinsella aerofaciens* sample and we confirmed the specificity of the PCR product by assaying also a DNA sample extracted from *Odoribacter splanchnicus* (ATCC cat# 29572). We also verified that both samples contained DNA from the 16S region by performing an additional PCR on both samples with the prokaryotic 16S rDNA universal primers at an annealing temperature of 55°C, forward (27F): 5'-AGAGTTTGATCCTGGCTCAG-3', reverse (1492R): 5'-CGGTTACCTTGTACGACTT-3' [10].

Cell culturing conditions for this experiment were the same as described above. On the day of treatment, the microbiota sample derived from individual 4 (chosen because this individual did not have a detectable amount of *Collinsella* at baseline or at any treatment time point) and the *Collinsella aerofaciens* was assessed by spectrophotometer (OD600) (Bio-Rad Smartspec 3000). Using these values, solutions were made with the microbiota sample at 10:1 to the number of colonocytes and the *Collinsella aerofaciens* was spiked into the microbiota sample at 4 dilutions: 10%, 1%, 0.1% and 0.01%. Additional wells containing only colonocytes and colonocytes with the microbiota sample (0% *Collinsella aerofaciens*) were cultured as controls on the same 12-well plate. Each treatment was performed in duplicate.

Following 2 hours of culturing, the wells were scraped on ice, pelleted and washed with cold PBS, resuspended in lysis buffer (Dynabeads mRNA Direct Kit) and stored at -80°C until extraction of colonocyte RNA (as described above).

## RNA-library preparation from colonocytes

Poly-adenylated mRNAs were isolated from thawed cell lysates using the Dynabeads mRNA Direct Kit (Ambion) and following the manufacturer’s instructions. RNA-seq libraries were prepared using a protocol modified from the NEBNext Ultradirectional (NEB) library preparation protocol to use Barcodes from BIOOScientific added by ligation, as described in [14]. The individual libraries were quantified using the KAPA real-time PCR system, following the manufacturer’s instructions and using a custom-made series of standards obtained from serial dilutions of the phi-X DNA (Illumina). The libraries were then pooled and sequenced on two lanes of the Illumina Next-seq 500 in the Luca/Pique laboratory using the high output kits for 75 cycles to obtain paired-end reads for an average of over 40 million total reads per sample.

## RNA sequencing and Alignment

Reads were aligned to the hg19 human reference genome using STAR [7] (<https://github.com/alexdobin/STAR/releases>, version STAR\_2.4.0h1), and the Ensemble reference transcriptome (version 75) with the following options:

```
STAR --runThreadN 12 --genomeDir <genome>
      --readFilesIn <fastqs.gz> --readFilesCommand zcat
      --outFileNamePrefix <stem> --outSAMtype BAM Unsorted
      --genomeLoad LoadAndKeep
```

where **<genome>** represents the location of the genome and index files, **<fastqs.gz>** represents that sample’s fastq files, and **<stem>** represents the filename stem of that sample. We further required a quality score of 10 to remove reads mapping to multiple locations.

## Differential Gene Expression Analysis

To identify differentially expressed (DE) genes, we used DESeq2 [12] (R version 3.2.1, DESeq2 version 1.8.1). Gene annotations from Ensembl version 75 were used and transcripts with fewer than 20 reads

total were discarded. `coverageBed` was utilized to count reads with `-s` to account for strandedness and `-split` for BED12 input. The counts were then utilized in DESeq2 with several models to determine changes in gene expression under various conditions. A gene was considered DE if at least one of its transcripts was DE. In order to identify genes that changed at each time point following co-culturing, we used each microbiota treatment as a replicate with the following model:

$$\begin{aligned} &\text{Gene expression} \sim \text{time point} + \text{treatment} \\ Y_{jn} &= \sum_t \tau_{jt} T_{tn} + \beta_{jt}^M M_{tn} \end{aligned} \tag{1}$$

where  $Y_{jn}$  represents the internal DEseq mean gene-expression parameter for gene  $j$  and experiment  $n$ ,  $T_{tn} = 1$  is the time-point indicator  $\{1, 2 \text{ and } 4\}$ ,  $M_n$  is the treatment indicator (control or microbiome), and  $\tau_{jt}$  and  $\beta_{jt}^M$  parameters are time-point specific baseline and microbiome effect respectively. With this model, we identified 1,835 genes that change after 1 hour (70% of genes increase in expression), 4,099 genes after 2 hours (53% of genes increase in expression) and 1,197 genes after 4 hours (56% increase) with BH FDR < 10%,  $|\log\text{FC}| > 0.25$  (Figure S2).

In order to identify genes that were differentially expressed at a given time point after co-culturing with a specific microbiota sample, we used the following model:

$$\begin{aligned} &\text{Gene expression} \sim \text{time point} + \text{microbiota sample} \\ Y_{jn} &= \sum_t \tau_{jt} T_{tn} + \sum_{tm} \beta_{jtm}^M M_{tmn} \end{aligned} \tag{2}$$

Compared to the first model, here we investigate the effect of each microbiome individual  $m \in \{1, 2, 3, 4, 5\}$  and time-point. Note that this model allows for a different effect to each microbiome compared to the previous model. With this model 1,131 genes changed after 1 hour with any of the 5 samples, 3,240 after 2 hours and 1,060 after 4 hours with BH FDR < 10%,  $|\log\text{FC}| > 0.25$  (Figure 1B).

We next used the likelihood ratio test that is a part of DESeq2 to compare the 2 models above in order to identify genes whose expression changes over time are determined by the individual from which the microbiota sample was taken. In this way, we identified 409 genes at BH FDR < 10%.

We considered a model utilizing baseline alpha diversity to determine if diversity itself, and not necessarily any component of the microbiota, influenced host cell gene expression:

$$\begin{aligned} &\text{Gene expression} \sim \text{time point} + \text{treatment} + \text{baseline alpha diversity} \\ Y_{jn} &= \sum_t \tau_{jt} T_{tn} + \beta_j^M M_n + \sum_m \alpha_j A_m M_{mn} \end{aligned} \quad (3)$$

Compared to the first model, we investigate the role of alpha diversity in host gene expression where  $\alpha_j$  is the alpha diversity-specific parameter and  $A_m$  represents the alpha diversity for each microbiota  $m$  sample at baseline. Baseline alpha diversity was measured with three metrics using QIIME: Chao1, Simpson index, and Shannon index. We identified 14, 53, and 7 genes that were associated with Chao1, Simpson index, and Shannon index, respectively. However, there were no genes that were differentially expressed in all three.

In order to identify components of the microbiota samples that affect gene expression we used the following model for each gene  $j$  and taxon  $g$ :

$$\begin{aligned} &\text{Gene expression} \sim \text{time point} + \text{treatment} + \text{baseline abundance} \\ Y_{jn} &= \sum_t \tau_{jt} T_{tn} + \beta_j^M M_n + \sum_m \gamma_{jg} G_{gm} M_{mn} \end{aligned} \quad (4)$$

Compared to the first model, we investigate the role of the baseline abundance of a given taxon in host gene expression where  $\gamma_{jg}$  is the taxon-specific parameter and  $G_{gm}$  represents the baseline abundance each microbiota sample  $m$  and each taxon  $g$  at baseline. Baseline abundance is the number of reads (after all samples have been rarified to the sample with the lowest read count of 141,000) for a given taxon in each of the uncultured samples. Each of the time points had the same baseline abundance. This model was run for 62 taxa that had at least 141 reads (0.1% of the total reads in a sample) in at least one of the five uncultured samples. Comparing each taxon to all genes expressed in the colonocytes, we had 9,125,927 tests. We identified 588 significant comparisons (BH FDR < 10%) comprising of 46 taxa and 121 genes (Table S6).

Finally, we analyzed the validation experiment with the spike-in of *Collinsella aerofaciens*. In order to identify genes that were differentially expressed because of the *Collinsella aerofaciens*, we used the following model:

$$\begin{aligned} \text{Gene expression} &\sim \text{treatment with microbiome} + \text{relative abundance of } \textit{Collinsella aerofaciens} \\ Y_{jn} &= \mu_j + \beta_j^M M_n + \gamma_j^C G_n^C M_n \end{aligned} \tag{5}$$

We investigate the role of the changing *Collinsella aerofaciens* concentration within the microbiota sample in host gene expression where  $\mu_j$  is the baseline mean,  $\beta_j^M$  is the overall microbiome effect, and  $\gamma_j^C$  is the *Collinsella*-specific parameter and  $G_n^C$  represents the spiked-in abundance *Collinsella*. Where treatment with microbiome is a factor with values yes and no denoting treatment with the overall microbiome sample and relative abundance of *Collinsella aerofaciens* is a number with the values 0, 0.01, 0.1, 1 and 10 depending on the spike-in amount of each sample. We identified 1,570 genes that change expression (BH FDR = 10%, Table S8, Figure S6) depending on the abundance of *Collinsella aerofaciens*. Note that this analysis has more power than the one across the five microbiomes.

## Gene ontology analysis

We utilized GeneTrail [1] to find enrichment of gene ontology terms. We compiled a list of unique genes that changed gene expression at any of the three time points (1hr, 2hr and 4hr) and determined which GO categories were under/over-represented as compared to a list of all genes expressed in colonocytes (28,107 genes). We considered a category over/under-represented if the Benjamini-Hochberg adjusted  $p$ -value  $< 0.05$  (Table S2, Figure S3).

## Enrichment of DE genes among genome-wide association studies

We downloaded the GWAS catalog [18] (version 1.0.1) on January 5th, 2016. To identify the overlap between DE genes in our dataset and those associated with a GWAS trait, we intersected genes that contain transcripts that change significantly under a given condition with the reported genes from the

GWAS catalog. We used a Fisher’s exact test on a 2x2 contingency table using 2 groups: genes that contain transcripts that are DE (“DEG”) and other genes that are expressed in each sample (“NOT”). We then split these groups into 2: genes that are associated with any GWAS trait (“TRAIT”) and genes that are not associated with any trait in the GWAS catalog (“NOT TRAIT”). Values are shown in Table S7.

When considering specific trait enrichment among the 1,570 genes differentially expressed in the spike-in validation experiment, 35 traits were tested. These included traits that had at least 5 differentially expressed genes and 1 non-differentially expressed gene associated with it.

## **Heatmap and Network of Gene Expression and Microbial Abundance**

Figure 3A was made by including all microbial taxa and host transcripts that were among the 588 transcript-by-taxon pairs with significant association. These were then filtered to only include transcripts that were significantly differentially expressed in at least 1 sample when each treatment is treated separately, leaving 393 transcript-by-taxon pairs including 219 transcripts (70 genes) and 28 taxa. The Spearman’s  $\rho$  was then calculated for all pairs of transcripts and taxa in the heatmap (including those that did not have significant association from the taxon-based model of DESeq2).

Figure 3B was made using the Spearman’s  $\rho$  of the same 393 transcript-by-taxon pairs. Transcripts with the lowest  $p$ -value (from correlation assessed by Spearman’s  $\rho$ ) were used to represent genes when multiple transcripts from the same gene had a significant association with a given microbial taxon (BH FDR = 10% from taxon-based model).

## **16S rRNA gene sequencing and analysis of the microbiome**

Half of each culturing well and the full volume of wells with microbiota samples cultured alone were used for extraction of microbial DNA using the PowerSoil kit from MO BIO Laboratories as directed, with a few modifications. Microbial DNA was also extracted from the uncultured microbial samples. Briefly, the sample was spun to collect live microbes. The pellet was then resuspended in 200 $\mu$ L of phenol:chloroform



and added to the 750 $\mu$ L bead solution from the PowerSoil kit. The kit protocol was then followed and the column was eluted in 60 $\mu$ L.

16S rRNA gene amplification and sequencing was performed at the University of Minnesota Genomics Center (UMGC), as described in Burns et al. [2]. Briefly, DNA isolated from the fecal microbiota was quantified by qPCR, and the V5-V6 regions of the 16S rRNA gene were PCR amplified with forward primer 5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG**RRGGATTAGATACCC**-3', and reverse primer 5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG**CGACRRCCATGCANCACT**-3', where the portion in bold is 16S-specific. Nextera indexing primers were added in the next PCR using the V5F primer 5'-AATGATACGGCGACCACCGAGATCTACAC[i5]TCGTCGGCAGCGTC-3', and V6R 5'-CAAGCAGAAGACGGCATAACGAGAT[i7]GTCTCGTGGGCTCGG-3', where [i5] and [i7] refer to the index sequences used by Illumina. This PCR was carried out using the KAPA HiFidelity Hot Start polymerase (Kapa Biosystems) for 20 cycles. The amplicons were then diluted 1:100 and used as input for a second PCR using different combinations of forward and reverse indexing primers for another 10 cycles. The pooled, size-selected product was diluted to 8pM, spiked with 15% PhiX and loaded onto an Illumina MiSeq instrument to generate the 16S rRNA gene sequences. Samples were heat denatured at 96 °C for 2 minutes prior to loading, and a MiSeq 600 cycle v3 kit was used to sequence the sample, resulting in 2.2 million raw reads per sample, on average. Barcodes were removed from the sample reads by UMGc and the Nextera adaptors were trimmed using CutAdapt 1.8.1.

The trimmed 16S rRNA gene sequence pairs were quality filtered (q-score  $\geq$  20, using QIIME 1.8.0) resulting in 1.41, 1.06, and 1.53 million high quality reads for sample replicates 1, 2, and 3, respectively [3, 15]. OTUs were picked using the closed reference algorithm against the Greengenes database (August, 2013 release) [2, 3, 6, 15]. The resulting OTU table was analyzed to determine microbial community diversity using QIIME scripts and rarefying to 141,000 reads. Rarefaction plot was generated using QIIME on the five uncultured samples (Figure S7B).

We verified that the fecal samples we utilized were similar to other healthy samples by comparing the

OTUs detected to the Human Microbiome Project data [9, 13]. 16S V4 OTU and HMP V1V3 OTU tables (<https://www.hmpdacc.org/HMQCP/>, final OTU table) were run through QIIME's `summarize_taxa.py` and consolidated at the L3 class level. Tables were merged by most prevalent classes (13) with the remaining taxa merged into other representing  $< 1\%$  of the total count on average. A Bray Curtis distance matrix and PCoA were performed in the R `vegan` package and plots were created with `ggplot2`, `grid` and `gridextra` libraries (Figure S4). HMP fecal samples colored red, study samples colored yellow and all other HMP samples colored grey.

## Determining Effect of Colonocytes on Microbiota Composition

OTUs were collapsed to the genera level using scripts in QIIME 1.9.1. In total, 292 taxa were detected across all samples and treatments. To filter the number of taxa, the relative abundances were summed for each taxon, and taxa with summed relative abundance greater than 0.003 were kept in the linear model. This yields 85 taxa of interest. To account for taxa that may be abundant at only at a particular time point, relative abundances of taxa were summed at each time point and kept if they had a total relative abundance greater than 0.0005. For time point 0, this threshold was 0.00025. The additional step yields 27 taxa; these were added onto the list of 85 for a total of 112 taxa.

To identify taxa that change significantly over time we used a linear model. Relative abundances of the 112 taxa of interest were normalized using an arcsine-square root transformation. In the linear model, “treatment” is binary and refers to microbiome samples grown with colonocytes or without colonocytes. In the model, this is treated as a factor. “Individual” refers to the individual microbiome being studied and is treated as a factor. Time is treated as a factor variable, with possible values of 1, 2 and 4 hours, in order to allow for changes that are not necessarily linear over time.

To assess how each taxon changed in response to culturing with colonocytes, we ran two linear models and compared the goodness of fit using a likelihood ratio test.  $P$ -values from the likelihood ratio test were extracted and adjusted using the Benjamini-Hochberg procedure.

Null Model

Transformed Relative Abundance of Taxa  $\sim$  Individual + Time

$$G_{gn} = \sum_t \tau_{gt} T_{tn} + \sum_m \beta_{gm}^M M_{mn} \quad (6)$$

Alternative Model

Transformed Relative Abundance of Taxa  $\sim$  Individual + Treatment + Time

$$G_{gn} = \sum_t \tau_{gt} T_{tn} + \sum_m \beta_{gm}^M M_{mn} + \gamma_g K_n \quad (7)$$

where  $G_{gn}$  represents the transformed relative abundance of taxa parameter for taxon  $g$  and experiment  $n$ ,  $T_{tn} = 1$  is the time-point indicator  $\{1, 2, 4\}$ ,  $M_{mn}$  is the treatment indicator  $m \in \{1, 2, 3, 4, 5\}$ , and  $\tau_{gt}$  and  $\beta_{gm}^M$  parameters are time-point specific baseline and individual microbiome effect respectively. The alternative model includes  $K_n$  as an indicator of the presence of colonocytes,  $\gamma_g$  is the colonocyte effect on each taxon. The model yielded 13/112 taxa that change significantly due to treatment with a BH FDR  $< 10\%$  (Table S5).

## Accession numbers for sequencing data

Submission of 16S rRNA sequencing data and RNA sequencing data of colonocytes in all conditions to Short Read Archive (SRA) is pending.

## Characteristics of the Gut Microbiome

Thirty-five microbial samples were assessed with 16S rRNA gene sequencing, including the five uncultured microbiota samples, each of these five cultured alone (without colonocytes) for 1, 2, and 4 hours, and co-cultured with colonocytes for 1, 2 and 4 hours. In total, there were 7 samples from each original microbiome, and 35 samples overall. QIIME was used to analyze the resulting sequences [3]. We found that the relative abundances of each taxa were most strongly affected by the individual from which the sample was derived and did not change dramatically over time (Figure S7A and C). The microbiome maintains its composition during culture and our sequencing depth is able to capture this variation (Figure S7B). These results demonstrate that the gut microbiota samples are typical and unaffected by culturing for up to 4 hours.

## Response of Human Colonocytes to Inoculation with Microbiome

When we consider how the microbiota influences the host cell response, we identify 3,904 genes that are differentially expressed in at least one sample at 1, 2 or 4 hours compared to control colonocytes. In order to gain power to identify shared differentially expressed genes, we next utilized a model where treatments at a given time point are considered replicates and compared to control colonocyte samples. In this model, we identified 4,099 genes (BH FDR < 10%,  $|\log_2 \text{FC}| > 0.25$ ) that change expression consistently across the five microbiota treatments at 2 hours, while over 1,000 change consistently at 1 and 4 hours (examples in Figure 1D, Figure S2A and B, Table S1). We were unable to identify any genes that were differentially expressed based on diversity of the microbiota sample at baseline, independent of diversity measure. However, we identified 121 genes that were associated with microbial taxa at baseline.

## Effect of Human Colonocytes on Microbiome Composition

We assessed the microbial composition of the individual microbiome samples grown with and without colonocytes. Microbiome samples grown with colonocytes were collected concurrently and from the same

wells that RNA-seq data was generated. We identified 13 taxa that showed varying abundance dependent on the presence of host cells (Figure S5, Table S5). These data demonstrate that interaction with the host affects microbial communities. We co-culture the colonocytes and microbiota for up to four hours, but it is likely that in vivo, the various species will eventually come to an equilibrium which can then be maintained in the host organism and will reflect both host genetic variation and diet. When this equilibrium is disrupted, previous studies have shown that disease and infection may occur [4, 5, 8, 11, 17].

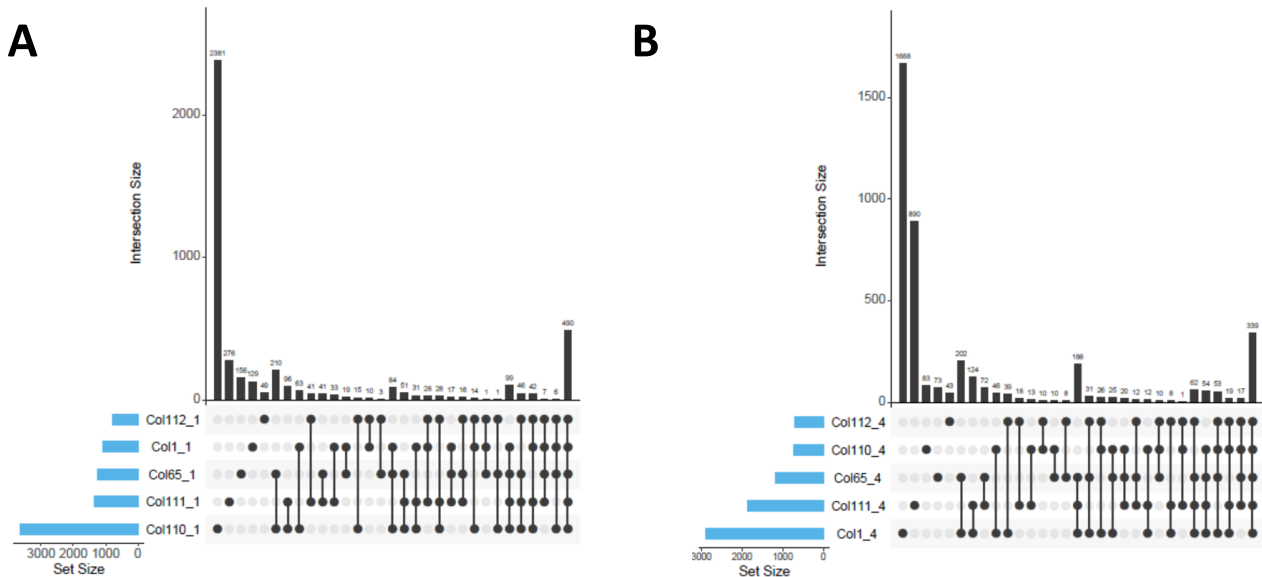
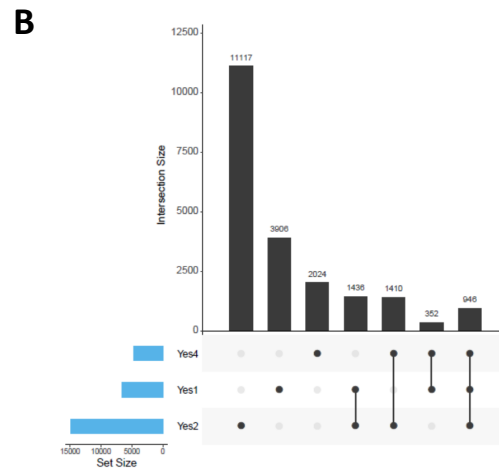
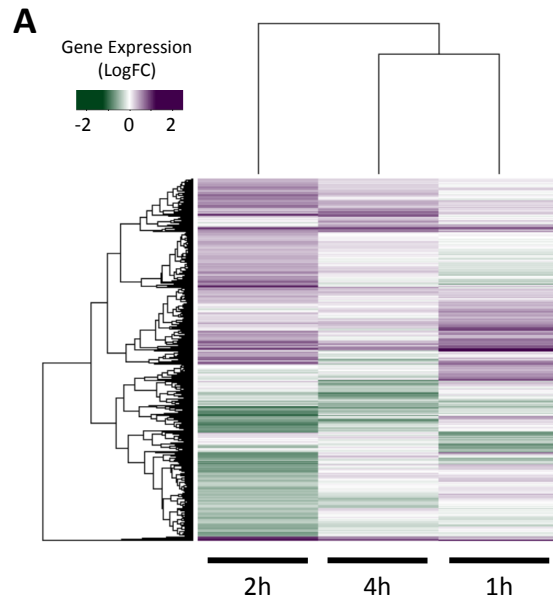


Figure S1: **Comparison of Differentially Expressed Genes Across Time Points and Microbiota.** **A** These plots were generated by UpSetR package to show the comparison of genes that are differentially expressed at 1 hour following exposure with each of the five microbiota. **B** show the comparison of differentially expressed genes across the five treatments at 4hr.



**Figure S2: Differentially Expressed Genes Across Time Points.** **A** This heatmap shows gene expression changes from the model in which the five microbiota treatments at a given time point are used as replicates. Purple denotes an increase in gene expression (green shows a reduction) compared to the gene expression in colonocytes cultured alone. Only genes that are differentially expressed in at least one sample are shown here. **B** Plot was generated by UpSetR package to show the comparison of genes across time points in the model where the five treatments were treated as replicates. The blue bars to the left show the total number of differentially expressed genes in the given set. The gray vertical bars show the number of genes that are in the set denoted below them. Sets with a single dark gray circle show the number of differentially expressed genes unique so that sample.

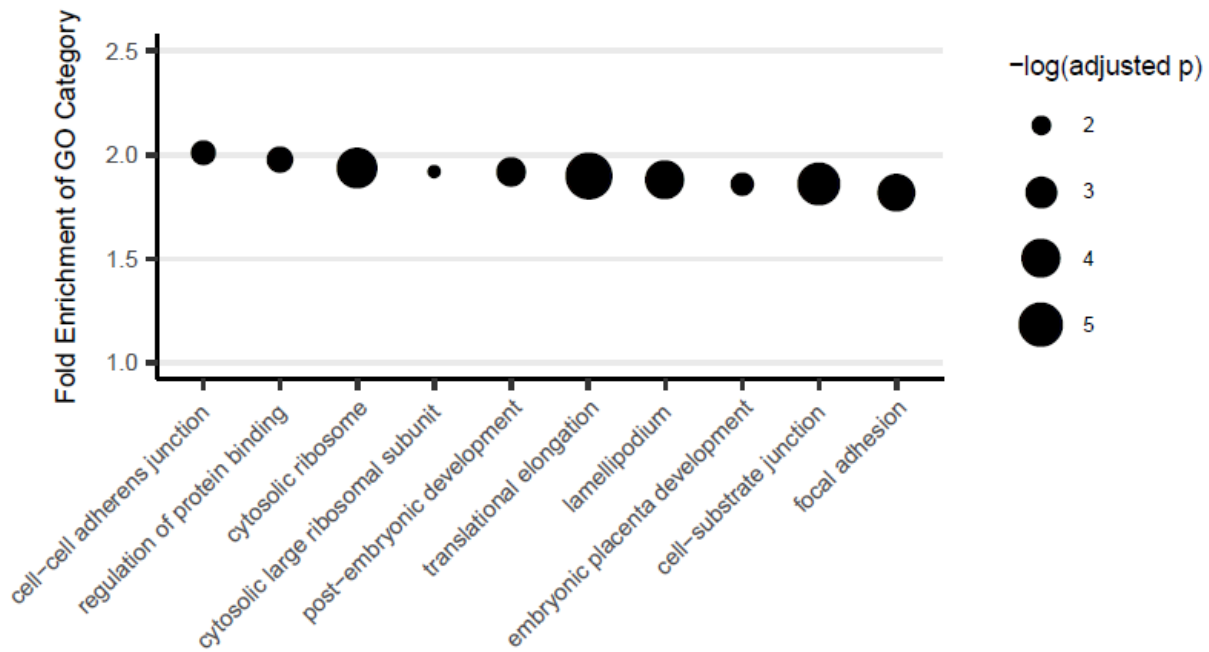


Figure S3: **Gene Ontology for DEG at any Time Point** Plot showing gene ontology enrichment for genes that were differentially expressed at any time point in the model in which each treatment is considered a replicate (and treatments lead to similar expression changes). The top 10 enrichments are shown in this plot, excluding categories where the expected number of genes is less than 10 or greater than 500.



Table S1: **Differentially Expressed Genes at Each Time Point.** Table denotes all transcripts that are differentially expressed in at least one time point. Genes are considered differentially expressed if at least one transcript is differentially expressed.

<http://genome.grid.wayne.edu/5Micro/TableS1.txt>

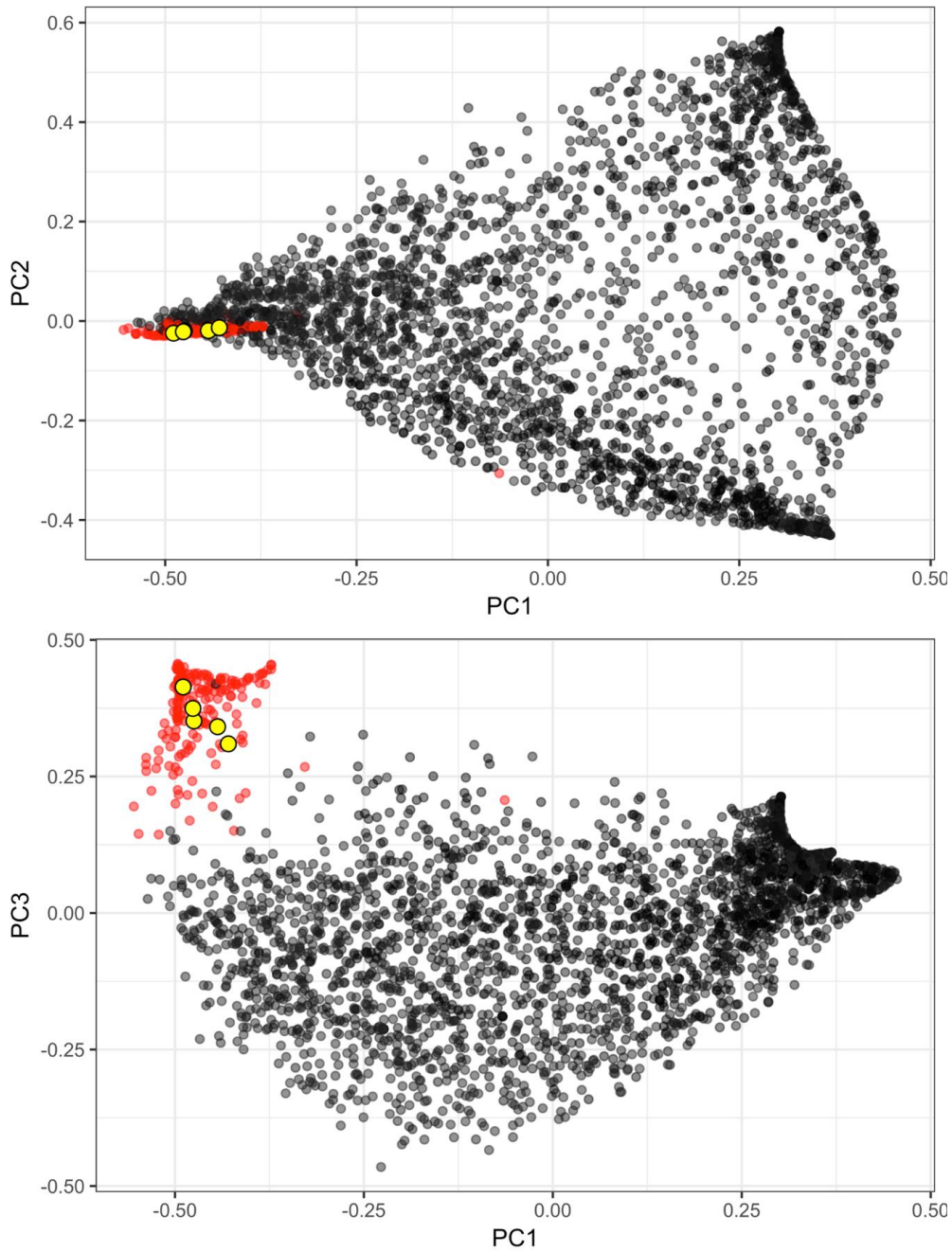


Figure S4: **Comparison of Fecal Samples to Human Microbiome Project.** PCoA representing samples from HMP and the current study, with HMP fecal samples in red and all other HMP samples in grey. The five fecal samples used in the current study are colored yellow. Top panel shows PC1 (x-axis) vs PC2 (y-axis), while bottom panel shows PC1 (x-axis) vs PC3 (y-axis).

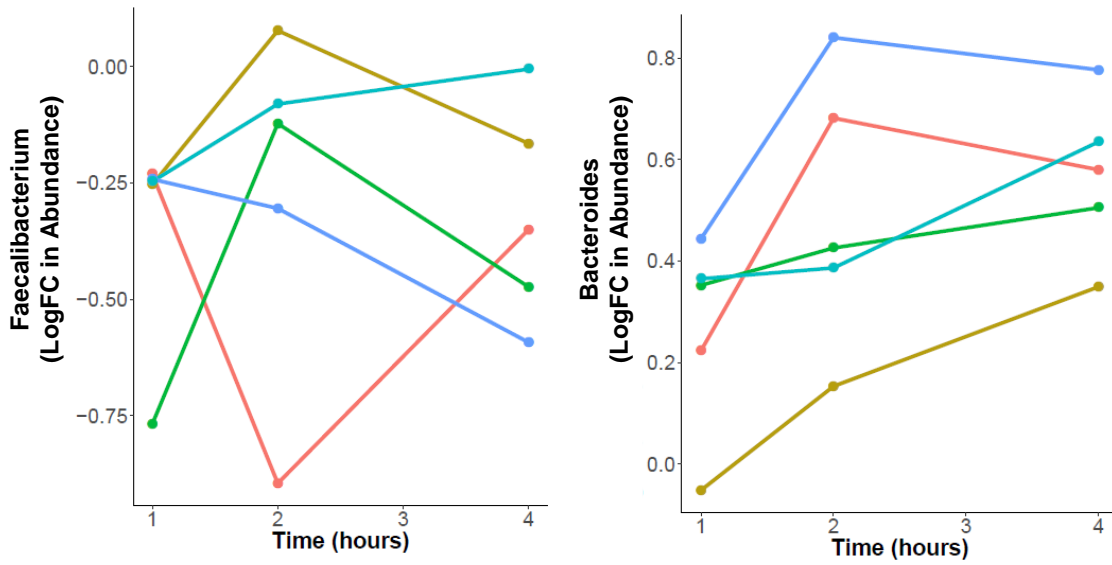


Figure S5: **Examples of Taxa Affected by Culturing with Colonocytes.** Genera (*Faecalibacterium* and *Bacteroides*) whose changes in abundance following co-culturing is inconsistent dependent on the microbiota sample in which it is found.

Table S2: **Gene Ontology Enrichment for DEG Across Time Points.** This table includes the Gene Ontology results from GeneTrail for adjusted  $p$ -value  $< 0.05$ . The genes tested were those that were differentially expressed in at least one time point when the five treatments were treated as replicates in the model. The background included all genes expressed in the colonocytes at all time points.

GO Category	Expected	Observed	p-value (fdr)
intracellular	2901.8	3139	2.40E-18
intracellular part	2815.44	3048	4.03E-17
cytoplasm	2022.98	2262	1.21E-15
protein binding	2104.28	2337	6.89E-15
regulation of cellular process	1562.88	1786	1.66E-14
positive regulation of cellular process	511.33	660	1.40E-13
regulation of biological process	1638.22	1853	2.90E-13
positive regulation of biological process	556.894	704	1.97E-12
biological regulation	1735.01	1940	7.09E-12
cytosol	363.619	477	1.44E-10
binding	3084.95	3255	1.64E-10
non-membrane-bounded organelle	696.564	843	2.33E-10
intracellular non-membrane-bounded organelle	696.564	843	2.33E-10
intracellular signaling pathway	392.804	508	2.57E-10
cytoskeleton	350.813	460	2.97E-10
positive regulation of biosynthetic process	178.385	255	2.91E-09
positive regulation of cellular biosynthetic process	176.002	252	2.98E-09
cellular process	2933.97	3099	3.52E-09
cellular protein metabolic process	662.614	793	1.47E-08
negative regulation of cellular process	480.656	593	2.79E-08
positive regulation of metabolic process	258.792	344	3.10E-08
positive regulation of macromolecule biosynthetic process	168.855	239	3.10E-08
intracellular signal transduction	306.738	398	3.88E-08
signal transduction	558.383	676	4.23E-08
positive regulation of cellular metabolic process	245.688	328	4.23E-08
signaling pathway	629.558	752	5.67E-08
regulation of signaling pathway	263.259	346	1.05E-07
intracellular organelle	2400.6	2563	1.34E-07
organelle	2403.28	2564	1.90E-07
positive regulation of macromolecule metabolic process	243.604	322	1.91E-07

anatomical structure morphogenesis	338.305	429	1.95E-07
positive regulation of transcription, DNA-dependent	124.78	182	2.26E-07
positive regulation of RNA metabolic process	126.567	184	2.44E-07
actin cytoskeleton	73.5576	118	2.55E-07
positive regulation of nitrogen compound metabolic process	170.94	236	3.89E-07
anchoring junction	45.2662	80	4.42E-07
protein modification process	480.06	582	5.60E-07
adherens junction	40.5014	73	6.73E-07
cytoskeletal protein binding	137.585	195	7.99E-07
cell	3608.49	3713	8.81E-07
cell part	3608.49	3713	8.81E-07
cellular component movement	159.921	221	9.68E-07
negative regulation of biological process	522.646	626	1.01E-06
positive regulation of transcription	145.924	204	1.20E-06
actin binding	85.7676	131	1.23E-06
nucleus	1341.91	1485	1.23E-06
positive regulation of gene expression	153.667	213	1.23E-06
cell leading edge	45.564	79	1.31E-06
regulation of cell proliferation	208.165	276	1.42E-06
negative regulation of macromolecule metabolic process	219.779	289	1.62E-06
cellular metabolic process	1917.86	2068	1.71E-06
regulation of metabolic process	989.305	1118	1.83E-06
cellular component organization	731.705	847	1.87E-06
regulation of transcription from RNA polymerase II promoter	189.701	253	2.93E-06
positive regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	165.579	225	2.93E-06
translational elongation	28.887	55	3.00E-06
cytoplasmic part	1376.15	1515	3.18E-06
actin filament-based process	76.8335	118	3.54E-06
signaling process	644.448	751	4.09E-06
organ development	462.192	555	4.13E-06
phosphorus metabolic process	346.346	428	4.34E-06
phosphate metabolic process	346.346	428	4.34E-06
negative regulation of macromolecule biosynthetic process	159.623	217	4.37E-06
cell proliferation	290.061	365	5.05E-06
signal transmission	642.661	748	5.05E-06
macromolecule modification	502.098	597	5.45E-06
regulation of cell communication	296.613	372	5.51E-06
actin cytoskeleton organization	73.5576	113	5.73E-06

regulation of cellular metabolic process	941.359	1062	5.73E-06
developmental process	825.811	941	5.73E-06
regulation of cellular component organization	145.924	200	6.42E-06
protein metabolic process	772.504	884	6.84E-06
signaling	879.118	996	6.84E-06
cytoskeletal part	234.967	302	7.04E-06
negative regulation of cellular metabolic process	214.121	278	8.15E-06
cellular macromolecule metabolic process	1483.06	1617	1.11E-05
organelle organization	399.653	483	1.12E-05
cytoskeleton organization	139.372	191	1.12E-05
negative regulation of metabolic process	235.265	301	1.12E-05
regulation of signaling process	206.08	268	1.12E-05
regulation of macromolecule metabolic process	859.463	973	1.12E-05
metabolic process	2140.91	2279	1.30E-05
phosphorylation	303.165	376	1.48E-05
cytosolic part	40.5014	69	1.52E-05
post-translational protein modification	409.183	492	1.60E-05
anatomical structure development	687.63	790	1.73E-05
cell cortex	33.9497	60	1.74E-05
regulation of signal transduction	204.591	265	1.75E-05
cellular macromolecule biosynthetic process	901.155	1014	1.83E-05
cell-substrate junction	26.8024	50	1.88E-05
macromolecule biosynthetic process	914.556	1027	2.24E-05
cellular biosynthetic process	1097.71	1217	2.30E-05
regulation of cellular component movement	61.9432	96	2.42E-05
apoptosis	294.826	365	2.48E-05
molecular_function	3687.11	3772	2.66E-05
negative regulation of transcription	133.416	182	2.66E-05
positive regulation of cellular component movement	35.1409	61	2.84E-05
negative regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	150.093	201	3.18E-05
regulation of primary metabolic process	898.177	1008	3.18E-05
programmed cell death	296.613	366	3.30E-05
negative regulation of cellular biosynthetic process	164.09	217	3.31E-05
negative regulation of biosynthetic process	166.175	219	3.92E-05
negative regulation of gene expression	149.795	200	4.14E-05
primary metabolic process	1956.87	2087	4.58E-05
cytosolic ribosome	21.1441	41	4.73E-05
nuclear lumen	420.797	500	5.27E-05

transcription from RNA polymerase II promoter	234.074	295	5.33E-05
vasculature development	86.661	125	5.44E-05
biosynthetic process	1124.81	1240	5.45E-05
negative regulation of nitrogen compound metabolic process	151.284	201	5.45E-05
regulation of cell migration	55.0938	86	5.58E-05
cell projection assembly	27.6958	50	5.77E-05
regulation of macromolecule biosynthetic process	720.984	819	6.05E-05
transcription factor binding	143.244	191	7.38E-05
regulation of gene expression	739.15	837	7.69E-05
multicellular organismal development	747.488	845	9.04E-05
regulation of protein metabolic process	175.109	227	9.04E-05
macromolecular complex	824.024	925	9.25E-05
lamellipodium	22.3353	42	9.47E-05
cell migration	111.379	153	0.000109135
regulation of cellular biosynthetic process	750.466	847	0.000112549
cell death	326.095	394	0.000118096
regulation of biosynthetic process	755.529	852	0.000118764
positive regulation of cell migration	32.7584	56	0.000121849
regulation of locomotion	60.1564	91	0.000130215
death	326.691	394	0.000139549
regulation of cellular protein metabolic process	154.263	202	0.000141101
nucleoside-triphosphatase regulator activity	118.824	161	0.000146372
focal adhesion	24.1221	44	0.000150645
enzyme regulator activity	224.544	281	0.000154948
regulation of molecular function	276.064	338	0.000157986
cell motility	119.122	161	0.000167541
localization of cell	119.122	161	0.000167541
tissue development	196.253	249	0.000170866
protein complex	682.269	773	0.000171576
macromolecule metabolic process	1614.69	1733	0.00017818
posttranscriptional regulation of gene expression	65.5169	97	0.000178367
cell-substrate adherens junction	25.0155	45	0.00018035
regulation of binding	58.3696	88	0.000199703
biological_process	3433.68	3525	0.000200174
platelet-derived growth factor receptor signaling pathway	5.95608	16	0.000203329
positive regulation of locomotion	34.8431	58	0.000209506
positive regulation of transcription from RNA polymerase II promoter	94.1061	131	0.000217224
blood vessel development	84.8742	120	0.000218296

negative regulation of transcription, DNA-dependent	106.912	146	0.000223501
organelle part	1337.14	1449	0.000225639
negative regulation of RNA metabolic process	108.698	148	0.000231044
GTPase regulator activity	116.441	157	0.000232979
in utero embryonic development	58.6674	88	0.000239779
protein amino acid phosphorylation	215.312	269	0.000250216
intracellular protein kinase cascade	156.645	203	0.000250216
signal transmission via phosphorylation event	156.645	203	0.000250216
regulation of apoptosis	233.478	289	0.000266288
system development	619.135	704	0.00026711
intracellular organelle part	1318.97	1429	0.000275945
transcription activator activity	113.463	153	0.000287888
transcription repressor activity	89.639	125	0.000291774
nucleotide binding	600.969	684	0.000315433
epithelium development	89.0434	124	0.000338421
regulation of programmed cell death	235.265	290	0.000359124
regulation of cell death	237.35	292	0.000396573
nuclear part	536.047	614	0.00041834
transcription regulator activity	366.895	433	0.000432651
ribonucleotide binding	493.164	568	0.000458054
purine ribonucleotide binding	493.164	568	0.000458054
actomyosin	9.23193	21	0.000492013
organ morphogenesis	154.858	199	0.000520193
negative regulation of transcription from RNA polymerase II promoter	75.6422	107	0.000575533
cell projection organization	120.908	160	0.000575533
regulation of actin cytoskeleton organization	23.8243	42	0.000588398
locomotion	142.946	185	0.000621588
basolateral plasma membrane	59.263	87	0.000649511
regulation of actin filament-based process	24.7177	43	0.000709148
small GTPase mediated signal transduction	117.037	155	0.000714964
embryonic development	155.752	199	0.000749912
immune system development	88.15	121	0.000859805
regulation of cellular component size	88.15	121	0.000859805
growth	131.629	171	0.000962241
ligase activity, forming carbon-nitrogen bonds	77.4291	108	0.00101729
regulation of nitrogen compound metabolic process	760.294	845	0.00102572
regulation of cellular component biogenesis	40.5014	63	0.00103057
regulation of organelle organization	65.8147	94	0.00107746



anatomical structure formation involved in morphogenesis	116.441	153	0.00122445
regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	754.636	838	0.00124298
multi-organism process	188.51	234	0.00133074
hemopoietic or lymphoid organ development	82.1939	113	0.00136475
purine nucleotide binding	514.605	585	0.00155803
chordate embryonic development	94.4039	127	0.00155803
embryonic development ending in birth or egg hatching	95.2973	128	0.00157203
positive regulation of cellular component organization	58.9652	85	0.00161657
post-embryonic development	14.5924	28	0.00182588
gene expression	971.437	1060	0.00195442
translation	112.272	147	0.00200423
protein C-terminus binding	41.3948	63	0.00215778
actin filament bundle	8.63632	19	0.00217762
actin filament organization	36.6299	57	0.00218073
stress fiber	8.04071	18	0.0023805
cell-cell junction	52.1157	76	0.0023805
ectoderm development	52.1157	76	0.0023805
response to chemical stimulus	368.086	427	0.00246874
intracellular receptor mediated signaling pathway	25.909	43	0.00258491
response to cytokine stimulus	29.7804	48	0.00262595
cellular developmental process	463.383	528	0.00272468
positive regulation of proteasomal ubiquitin-dependent protein catabolic process	6.25389	15	0.0027872
blood vessel morphogenesis	74.1532	102	0.0027872
cell communication	412.756	474	0.00283834
cell junction	136.096	173	0.00290292
enzyme linked receptor protein signaling pathway	118.526	153	0.00303749
small GTPase regulator activity	79.5137	108	0.00315522
positive regulation of cell death	126.567	162	0.00315522
protein domain specific binding	108.996	142	0.00318674
growth factor receptor binding	20.8463	36	0.00321904
regulation of anatomical structure size	103.04	135	0.00342881
regulation of cytoskeleton organization	37.2255	57	0.00346201
positive regulation of apoptosis	124.184	159	0.00349103
magnesium ion binding	46.1596	68	0.00351987
regulation of phosphate metabolic process	144.733	182	0.0036185
regulation of phosphorus metabolic process	144.733	182	0.0036185
regulation of catalytic activity	230.798	277	0.00364966

regulation of protein binding	11.6144	23	0.00374553
organelle lumen	518.179	584	0.00380067
regulation of transcription	657.849	730	0.0041035
regulation of intracellular protein kinase cascade	87.8522	117	0.00416823
transcription corepressor activity	39.9057	60	0.00417879
Ras protein signal transduction	63.1345	88	0.00432796
immune system process	238.839	285	0.0044506
hemopoiesis	77.7269	105	0.00460345
transcription	684.652	757	0.00476094
positive regulation of programmed cell death	125.078	159	0.00489865
cell-cell adherens junction	10.4231	21	0.0049525
ruffle	20.5485	35	0.00516833
regulation of translation	40.2036	60	0.00516833
negative regulation of cell proliferation	95.2973	125	0.00516833
regulation of transcription, DNA-dependent	436.581	496	0.00532992
intracellular organelle lumen	509.543	573	0.00532992
regulation of localization	181.065	221	0.00537076
membrane-enclosed lumen	527.709	592	0.00540218
I-kappaB kinase/NF-kappaB cascade	48.5421	70	0.00558999
cytosolic small ribosomal subunit	9.82753	20	0.00558999
positive regulation of cell proliferation	109.592	141	0.00571284
microtubule cytoskeleton	162.303	200	0.00573974
negative regulation of signaling pathway	68.7927	94	0.00573974
regulation of RNA metabolic process	450.28	510	0.00574012
ribosome binding	4.76487	12	0.00583793
regulation of DNA binding	46.1596	67	0.00584973
nuclear periphery	19.9529	34	0.00592617
regulation of phosphorylation	139.075	174	0.00595656
angiogenesis	59.5608	83	0.00596818
nucleolus	204.294	246	0.0059894
regulation of developmental process	199.827	241	0.00617473
embryonic placenta development	13.4012	25	0.0063011
acid-amino acid ligase activity	69.0905	94	0.0065973
double-stranded DNA binding	24.7177	40	0.00703082
ligase activity	120.015	152	0.00735058
transmembrane receptor protein tyrosine kinase signaling pathway	75.3444	101	0.00748567
ruffle membrane	8.04071	17	0.00812845
intracellular membrane-bounded organelle	2176.05	2268	0.00816714

membrane-bounded organelle	2178.14	2270	0.00822458
enzyme binding	176.3	214	0.00866638
positive regulation of binding	31.2694	48	0.00872093
response to stimulus	782.331	854	0.00905681
copper ion homeostasis	2.68024	8	0.00906384
epidermis development	48.5421	69	0.00920927
positive regulation of molecular function	162.005	198	0.00925994
regulation of anatomical structure morphogenesis	81.0027	107	0.00929841
cell differentiation	437.772	494	0.00955025
filopodium assembly	6.84949	15	0.0100589
muscle structure development	77.7269	103	0.0101753
cell adhesion molecule binding	9.52973	19	0.0102525
ATP binding	399.653	453	0.0109405
positive regulation of cell communication	106.018	135	0.0111426
kinase activity	213.823	254	0.0111426
structure-specific DNA binding	38.1189	56	0.0113348
protein targeting	76.2378	101	0.0113632
adenyl ribonucleotide binding	405.907	459	0.0122775
transferase activity, transferring phosphorus-containing groups	247.475	290	0.0123376
mRNA 3'-UTR binding	5.06267	12	0.0124833
epidermal growth factor receptor binding	3.87145	10	0.0124833
epithelial cell migration	3.87145	10	0.0124833
regulation of epithelial cell migration	3.87145	10	0.0124833
cellular response to chemical stimulus	115.25	145	0.0124833
tissue migration	3.87145	10	0.0124833
epithelium migration	3.87145	10	0.0124833
macromolecule localization	343.07	392	0.0129715
morphogenesis of an epithelium	55.0938	76	0.0131713
transcription, DNA-dependent	473.806	530	0.0132267
nucleus organization	8.33851	17	0.0132844
regulation of proteasomal ubiquitin-dependent protein catabolic process	8.33851	17	0.0132844
positive regulation of signaling pathway	98.5731	126	0.0132844
sequence-specific DNA binding	136.99	169	0.0132844
cytosolic large ribosomal subunit	10.4231	20	0.0135145
RNA biosynthetic process	474.998	531	0.013627
positive regulation of DNA binding	27.1002	42	0.013663
negative regulation of survival gene product expression	1.78682	6	0.0137852
nucleoplasm part	166.175	201	0.0138743

regulation of I-kappaB kinase/NF-kappaB cascade	36.9277	54	0.0149107
nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	1101.58	1178	0.0149614
regulation of Ras protein signal transduction	61.3476	83	0.0149734
placenta development	21.7397	35	0.0154884
phosphatase regulator activity	13.4012	24	0.0154884
regulation of cell adhesion	37.8211	55	0.0154884
extrinsic to membrane	24.1221	38	0.015929
protein serine/threonine kinase activity	119.717	149	0.0168327
localization	846.061	915	0.0168327
microspike assembly	7.1473	15	0.0170997
perinuclear region of cytoplasm	88.7456	114	0.0180444
regulation of actin polymerization or depolymerization	14.2946	25	0.0186624
response to organic substance	224.246	263	0.0186624
small conjugating protein ligase activity	59.263	80	0.0192379
smooth muscle cell migration	6.55169	14	0.0194021
anti-apoptosis	61.0498	82	0.0198108
negative regulation of protein metabolic process	61.0498	82	0.0198108
tissue morphogenesis	70.5796	93	0.0198407
interspecies interaction between organisms	90.8302	116	0.0203992
regulation of cell size	73.2598	96	0.0204132
GTPase activity	56.8806	77	0.0209859
negative regulation of kinase activity	30.0782	45	0.0209859
cellular protein localization	140.266	171	0.0209859
Rho protein signal transduction	33.3541	49	0.0210089
Rho GTPase binding	8.63632	17	0.0210089
purine nucleoside binding	435.687	487	0.0212106
regulation of receptor activity	5.95608	13	0.0213252
nucleoside binding	438.665	490	0.0216245
cellular copper ion homeostasis	2.38243	7	0.0219516
regulation of filopodium assembly	2.38243	7	0.0219516
positive regulation of filopodium assembly	2.38243	7	0.0219516
negative regulation of cellular component organization	47.6487	66	0.02223
receptor signaling protein activity	43.4794	61	0.0229058
negative regulation of transferase activity	31.865	47	0.0229833
regulation of cell morphogenesis	44.3728	62	0.0235721
response to mechanical stimulus	17.5704	29	0.0236476
nucleoplasm	260.876	301	0.0245209
cellular component biogenesis	304.951	348	0.0245209

cellular macromolecule localization	140.861	171	0.0247884
membrane organization	116.441	144	0.0248166
regulation of actin filament length	14.5924	25	0.0251586
positive regulation of I-kappaB kinase/NF-kappaB cascade	33.6519	49	0.0251586
cell adhesion mediated by integrin	4.76487	11	0.0251834
GTPase activator activity	61.6454	82	0.0255069
transcription factor activity	218.29	255	0.0257594
positive regulation of epithelial cell migration	2.97804	8	0.0257594
myeloid cell differentiation	38.7145	55	0.0263875
retinoic acid receptor signaling pathway	4.16926	10	0.0265416
adenyl nucleotide binding	426.753	476	0.0271917
muscle cell migration	7.4451	15	0.0274236
response to lipopolysaccharide	28.887	43	0.0274236
regulation of Rho protein signal transduction	28.887	43	0.0274236
regulation of hydrolase activity	89.9368	114	0.0274236
regulation of transferase activity	108.698	135	0.0274236
regulation of biological quality	417.521	466	0.0282081
organelle assembly	15.4858	26	0.0283474
phosphotransferase activity, alcohol group as acceptor	197.444	232	0.0283863
actin polymerization or depolymerization	19.3573	31	0.0284777
protein phosphatase regulator activity	12.5078	22	0.0290744
cellular membrane organization	116.144	143	0.0294093
cell morphogenesis	110.783	137	0.0297853
negative regulation of organelle organization	27.398	41	0.0297853
negative regulation of cell communication	84.8742	108	0.0297853
nuclear matrix	17.8682	29	0.0299112
negative regulation of apoptosis	103.636	129	0.0299112
negative regulation of programmed cell death	104.529	130	0.0299112
negative regulation of molecular function	102.742	128	0.0299112
phospholipid binding	50.0311	68	0.030404
regulation of proteolysis	20.2507	32	0.0308853
intracellular protein transport	123.589	151	0.031015
cell cycle	279.34	319	0.0329926
specific transcriptional repressor activity	10.4231	19	0.033005
regulation of kinase activity	104.827	130	0.033005
cell growth	68.4949	89	0.0342581
cellular component assembly	269.215	308	0.0342581
negative regulation of cellular protein metabolic process	57.1784	76	0.0343739
homeostasis of number of cells	30.0782	44	0.0343739

catalytic activity	1361.26	1434	0.0350415
positive regulation of hydrolase activity	44.3728	61	0.0356181
entrainment of circadian clock	1.48902	5	0.0356283
actin-dependent ATPase activity	1.48902	5	0.0356283
motile primary cilium	1.48902	5	0.0356283
cell-substrate adhesion	39.3101	55	0.0356283
toll-like receptor 4 signaling pathway	1.48902	5	0.0356283
embryonic placenta morphogenesis	1.48902	5	0.0356283
multicellular organismal process	993.177	1059	0.0361428
regulation of protein kinase activity	100.658	125	0.0361428
ubiquitin-protein ligase activity	53.9025	72	0.0369254
response to molecule of bacterial origin	31.865	46	0.0370181
negative regulation of protein kinase activity	28.5892	42	0.0375518
regulation of small GTPase mediated signal transduction	71.473	92	0.0388688
MAP kinase kinase kinase activity	5.65828	12	0.0390179
channel regulator activity	12.8056	22	0.0390179
protein ubiquitination involved in ubiquitin-dependent protein catabolic process	5.65828	12	0.0390179
negative regulation of cell death	106.316	131	0.0392421
transcription cofactor activity	100.062	124	0.039806
ribosomal subunit	33.6519	48	0.039806
structural constituent of ribosome	43.7772	60	0.0398438
transferase activity	455.342	503	0.0406306
positive regulation of cell cycle	19.0595	30	0.0415451
phosphoinositide binding	31.2694	45	0.042582
cellular component morphogenesis	121.206	147	0.0444184
intracellular transport	207.867	241	0.0444857
cortical cytoskeleton	9.23193	17	0.0445754
regulation of anti-apoptosis	10.7209	19	0.0461865
regulation of protein complex assembly	23.2287	35	0.0477704
regulation of protein modification process	101.551	125	0.0485977
inclusion body	4.46706	10	0.048932
guanyl nucleotide binding	97.0841	120	0.0489917
guanyl ribonucleotide binding	97.0841	120	0.0489917
regulation of catabolic process	74.7488	95	0.0491576

**Table S3: Differentially Expressed Genes at Each Time Point and Treatment.** Table denotes all transcripts that are differentially expressed in at least one time point, treatment combination. Genes are considered differentially expressed if at least one transcript is differentially expressed.

<http://genome.grid.wayne.edu/5Micro/TableS3.txt>

**Table S4: Genes Expressed Differentially Across Treatments From Likelihood Ratio Test.** Table denotes all transcripts that are differentially expressed between the five microbiota treatments as determined by the likelihood ratio test in the DESeq2 package. Genes are considered differentially expressed if at least one transcript is differentially expressed.

<http://genome.grid.wayne.edu/5Micro/TableS4.txt>



<b>Bacteria Taxa</b>	<b>p-value</b>	<b>BH adjusted p-value</b>
k__Archaea.p__Euryarchaeota.c__DSEG.o__DHVE3.f__g__.	1.55E-09	1.73E-07
k__Bacteria.p__Bacteroidetes.c__Bacteroidia.o__Bacteroidales.f__Prevotellaceae.g__Prevotella	2.59E-07	1.45E-05
k__Bacteria.p__Bacteroidetes.c__Bacteroidia.o__Bacteroidales.f__Bacteroidaceae.g__Bacteroides	1.61E-05	0.000480663
k__Bacteria.p__Cyanobacteria.c__Chloroplast.o__Streptophyta.f__g__	1.72E-05	0.000480663
k__Bacteria.p__Bacteroidetes.c__Bacteroidia.o__Bacteroidales.f__Porphyromonadaceae.g__Candidatus.Azobacteroides	0.000288621	0.006465103
k__Bacteria.p__Proteobacteria.c__Betaproteobacteria.o__Burkholderiales.f__Comamonadaceae.g__.	0.000960123	0.017922301
k__Bacteria.p__Firmicutes.c__Clostridia.o__Clostridiales.f__Clostridiaceae.g__	0.001734073	0.027745174
k__Bacteria.p__Actinobacteria.c__Actinobacteria.o__Bifidobacteriales.f__Bifidobacteriaceae.g__Bifidobacterium	0.003306716	0.046294026
k__Bacteria.p__Firmicutes.c__Clostridia.o__Clostridiales.f__Ruminococcaceae.g__Faecalibacterium	0.005076328	0.063172086
k__Bacteria.p__Bacteroidetes.c__Bacteroidia.o__Bacteroidales.f__Bacteroidaceae.g__5.7N15	0.006464096	0.065816246
k__Bacteria.p__Bacteroidetes.c__Cytophagia.o__Cytophagales.f__Cytophagaceae.g__	0.006431979	0.065816246
k__Bacteria.p__Proteobacteria.c__Alphaproteobacteria.o__Rhizobiales.f__Methylobacteriaceae.g__Methylobacterium	0.008861207	0.079584783
k__Bacteria.p__Verrucomicrobia.c__Verrucomicrobiae.o__Verrucomicrobiales.f__Verrucomicrobiaceae.g__Akkermansia	0.009237519	0.079584783

Table S5: **Taxa Whose Abundance Depends on Co-Culturing with Colonocytes** Table denotes 13 taxa whose abundance is affected by culturing with colonocytes using the likelihood ratio test between models with and without a treatment effect (BH FDR < 10%).

Table S6: **Differentially Expressed Genes By Baseline Abundance of Microbial Taxa.** Table denotes all transcripts that are differentially expressed relative to the mean-centered baseline abundance of denoted taxon. Genes are considered differentially expressed if at least one transcript is differentially expressed.

<http://genome.grid.wayne.edu/5Micro/TableS6.txt>

Genus	TRAIT	DEG; TRAIT	NOT; TRAIT	DEG; NOT TRAIT	NOT; NOT TRAIT	p-value	Odds Ratio
Thiothrix	All GWAS Traits	18	6326	17	21349	0.000328	3.572967
Collinsella	All GWAS Traits	15	6329	15	21350	0.001467	3.373112
Streptococcus	All GWAS Traits	15	6329	16	21348	0.001937	3.16205
Odoribacter	All GWAS Traits	19	6325	21	21345	0.00093	3.053145
Thiovirga	All GWAS Traits	24	6320	30	21334	0.000485	2.700387
Blautia	All GWAS Traits	13	6331	53	21312	0.6599	0.825702

Table S7: **Enrichment for GWAS Traits Among DEG from Taxon Model.** Table delineates the 2x2 table used for a Fisher's Exact Test to determine enrichment among DE genes (from the taxon model) for GWAS traits.

Table S8: **Differentially Expressed Genes Following *Collinsella* Spike-In.** Table denotes all transcripts that are differentially expressed following co-culture of colonocytes with microbiome plus varying concentrations of *Collinsella aerofaciens*. Genes are considered differentially expressed if at least one transcript is differentially expressed.

<http://genome.grid.wayne.edu/5Micro/TableS8.txt>

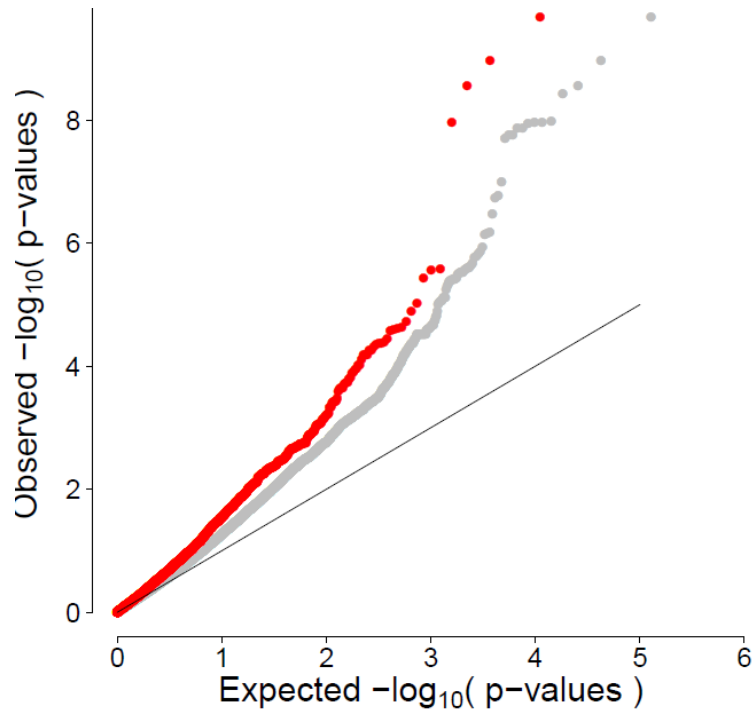


Figure S6: **Comparison of Taxon Model with *Collinsella* and Spike-In of *Collinsella aerofaciens*.** QQ-plot of  $p$ -values from the model of the five microbiota experiment with abundance of baseline *Collinsella* (gray). The red points indicate the same values from the gray but subsetting to only include 1,570 genes that are DEG in the validation spike-in experiment with *Collinsella aerofaciens*.

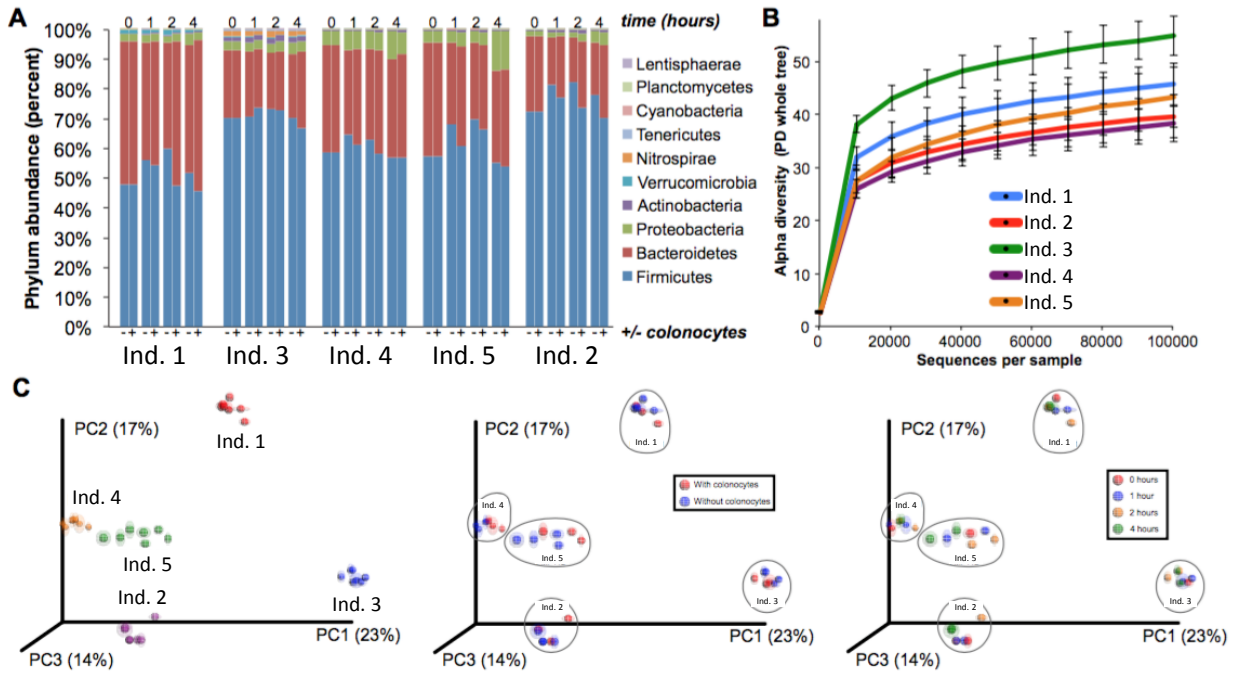


Figure S7: **Analysis of the Microbial Community Changes Over Time When Co-cultured with Normal Human Colonocytes.** **A** Relative abundances of the bacterial phyla in each of the five different microbial communities at 0 (uncultured), 1, 2 and 4 hours with and without co-culturing (+/-) with colonocytes. **B** Alpha diversity rarefaction plots for each of the microbial communities in this study. Error bars represent the standard deviation for the multiple rarefaction trials. **C** Principal coordinate analysis (unweighted unifracs) of the samples shows a clear separation by microbial community (left panel), while neither co-culture conditions (center panel) nor time (right panel) had a dramatic effect on the community structures.

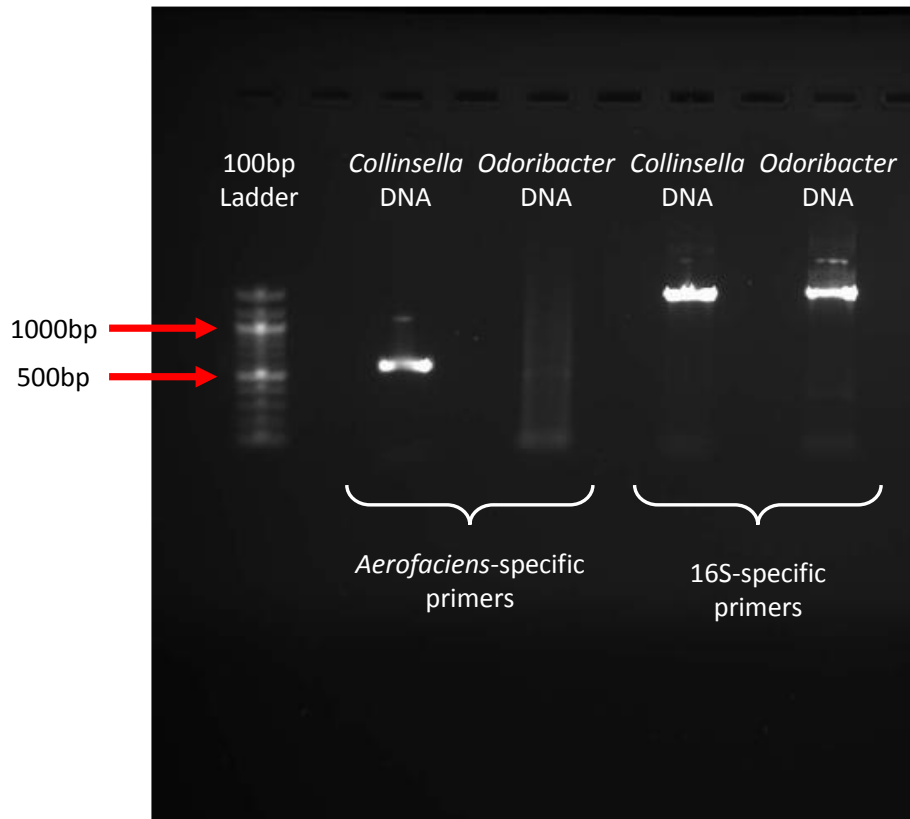


Figure S8: **Confirmation of *Collinsella aerofaciens* DNA.** Gel with five wells: 100bp ladder, *Collinsella aerofaciens* DNA and *Odoribacter splanchnicus* DNA with primers specific to 16S region of *Collinsella aerofaciens* (590bp), and *Collinsella aerofaciens* DNA and *Odoribacter splanchnicus* DNA with primers to amplify 16S region in any microbial sample.

## References

- [1] Backes, C., Keller, A., Kuentzer, J., Kneissl, B., Comtesse, N., Elnakady, Y. A., Muller, R., Meese, E., and Lenhof, H. P. (2007). GeneTrail–advanced gene set enrichment analysis. *Nucleic Acids Res.*, **35**(Web Server issue), W186–192.
- [2] Burns, M. B., Lynch, J., Starr, T. K., Knights, D., and Blekhman, R. (2015). Virulence genes are a signature of the microbiome in the colorectal tumor microenvironment. *Genome Med.*, **7**(1), 55.
- [3] Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Fierer, N., Pena, A. G., Goodrich, J. K., Gordon, J. I., Huttley, G. A., Kelley, S. T., Knights, D., Koenig, J. E., Ley, R. E., Lozupone, C. A., McDonald, D., Muegge, B. D., Pirrung, M., Reeder, J., Sevinsky, J. R., Turnbaugh, P. J., Walters, W. A., Widmann, J., Yatsunenko, T., Zaneveld, J., and Knight, R. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods*, **7**(5), 335–336.
- [4] Chassaing, B. and Darfeuille-Michaud, A. (2011). The commensal microbiota and enteropathogens in the pathogenesis of inflammatory bowel diseases. *Gastroenterology*, **140**(6), 1720–1728.
- [5] de Oliveira, G. L. V., Leite, A. Z., Higuchi, B. S., Gonzaga, M. I., and Mariano, V. S. (2017). Intestinal dysbiosis and probiotic applications in autoimmune diseases. *Immunology*.
- [6] DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., Huber, T., Dalevi, D., Hu, P., and Andersen, G. L. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.*, **72**(7), 5069–5072.
- [7] Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**(1), 15–21.
- [8] Hansen, R., Russell, R. K., Reiff, C., Louis, P., McIntosh, F., Berry, S. H., Mukhopadhyaya, I., Bisset, W. M., Barclay, A. R., Bishop, J., Flynn, D. M., McGrogan, P., Loganathan, S., Mahdi, G., Flint,



H. J., El-Omar, E. M., and Hold, G. L. (2012). Microbiota of de-novo pediatric IBD: increased Faecal-ibacterium prausnitzii and reduced bacterial diversity in Crohn's but not in ulcerative colitis. *Am. J. Gastroenterol.*, **107**(12), 1913–1922.

- [9] Huttenhower, C., Gevers, D., Knight, R., Abubucker, S., Badger, J. H., Chinwalla, A. T., Creasy, H. H., Earl, A. M., FitzGerald, M. G., Fulton, R. S., Giglio, M. G., Hallsworth-Pepin, K., Lobos, E. A., Madupu, R., Magrini, V., Martin, J. C., Mitreva, M., Muzny, D. M., Sodergren, E. J., Versalovic, J., Wollam, A. M., Worley, K. C., Wortman, J. R., Young, S. K., Zeng, Q., Aagaard, K. M., Abolude, O. O., Allen-Vercoe, E., Alm, E. J., Alvarado, L., Andersen, G. L., Anderson, S., Appelbaum, E., Arachchi, H. M., Armitage, G., Arze, C. A., Ayvaz, T., Baker, C. C., Begg, L., Belachew, T., Bhonagiri, V., Bihan, M., Blaser, M. J., Bloom, T., Bonazzi, V., Brooks, J., Buck, G. A., Buhay, C. J., Busam, D. A., Campbell, J. L., Canon, S. R., Cantarel, B. L., Chain, P. S., Chen, I. M., Chen, L., Chhibba, S., Chu, K., Ciulla, D. M., Clemente, J. C., Clifton, S. W., Conlan, S., Crabtree, J., Cutting, M. A., Davidovics, N. J., Davis, C. C., DeSantis, T. Z., Deal, C., Delehaunty, K. D., Dewhirst, F. E., Deych, E., Ding, Y., Dooling, D. J., Dugan, S. P., Dunne, W. M., Durkin, A., Edgar, R. C., Erlich, R. L., Farmer, C. N., Farrell, R. M., Faust, K., Feldgarden, M., Felix, V. M., Fisher, S., Fodor, A. A., Forney, L. J., Foster, L., Di Francesco, V., Friedman, J., Friedrich, D. C., Fronick, C. C., Fulton, L. L., Gao, H., Garcia, N., Giannoukos, G., Giblin, C., Giovanni, M. Y., Goldberg, J. M., Goll, J., Gonzalez, A., Griggs, A., Gujja, S., Haake, S. K., Haas, B. J., Hamilton, H. A., Harris, E. L., Hepburn, T. A., Herter, B., Hoffmann, D. E., Holder, M. E., Howarth, C., Huang, K. H., Huse, S. M., Izard, J., Jansson, J. K., Jiang, H., Jordan, C., Joshi, V., Katancik, J. A., Keitel, W. A., Kelley, S. T., Kells, C., King, N. B., Knights, D., Kong, H. H., Koren, O., Koren, S., Kota, K. C., Kovar, C. L., Kyrpides, N. C., La Rosa, P. S., Lee, S. L., Lemon, K. P., Lennon, N., Lewis, C. M., Lewis, L., Ley, R. E., Li, K., Liolios, K., Liu, B., Liu, Y., Lo, C. C., Lozupone, C. A., Lunsford, R., Madden, T., Mahurkar, A. A., Mannon, P. J., Mardis, E. R., Markowitz, V. M., Mavromatis, K., McCorrison, J. M., McDonald, D., McEwen, J., McGuire, A. L., McInnes, P., Mehta, T., Mihindukulasuriya, K. A., Miller, J. R., Minx, P. J., Newsham, I., Nusbaum, C.,

O’Laughlin, M., Orvis, J., Pagani, I., Palaniappan, K., Patel, S. M., Pearson, M., Peterson, J., Podar, M., Pohl, C., Pollard, K. S., Pop, M., Priest, M. E., Proctor, L. M., Qin, X., Raes, J., Ravel, J., Reid, J. G., Rho, M., Rhodes, R., Riehle, K. P., Rivera, M. C., Rodriguez-Mueller, B., Rogers, Y. H., Ross, M. C., Russ, C., Sanka, R. K., Sankar, P., Sathirapongsasuti, J., Schloss, J. A., Schloss, P. D., Schmidt, T. M., Scholz, M., Schriml, L., Schubert, A. M., Segata, N., Segre, J. A., Shannon, W. D., Sharp, R. R., Sharpton, T. J., Shenoy, N., Sheth, N. U., Simone, G. A., Singh, I., Smillie, C. S., Sobel, J. D., Sommer, D. D., Spicer, P., Sutton, G. G., Sykes, S. M., Tabbaa, D. G., Thiagarajan, M., Tomlinson, C. M., Torralba, M., Treangen, T. J., Truty, R. M., Vishnivetskaya, T. A., Walker, J., Wang, L., Wang, Z., Ward, D. V., Warren, W., Watson, M. A., Wellington, C., Wetterstrand, K. A., White, J. R., Wilczek-Boney, K., Wu, Y., Wylie, K. M., Wylie, T., Yandava, C., Ye, L., Ye, Y., Yooseph, S., Youmans, B. P., Zhang, L., Zhou, Y., Zhu, Y., Zoloth, L., Zucker, J. D., Birren, B. W., Gibbs, R. A., Highlander, S. K., Methe, B. A., Nelson, K. E., Petrosino, J. F., Weinstock, G. M., Wilson, R. K., and White, O. (2012). Structure, function and diversity of the healthy human microbiome. *Nature*, **486**(7402), 207–214.

- [10] Kageyama, A., Sakamoto, M., and Benno, Y. (2000). Rapid identification and quantification of *Collinsella aerofaciens* using PCR. *FEMS Microbiol. Lett.*, **183**(1), 43–47.
- [11] Lin, L. and Zhang, J. (2017). Role of intestinal microbiota and metabolites on gut homeostasis and human diseases. *BMC Immunol.*, **18**(1), 2.
- [12] Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, **15**(12), 550.
- [13] Methe, B. A., Nelson, K. E., Pop, M., Creasy, H. H., Giglio, M. G., Huttenhower, C., Gevers, D., Petrosino, J. F., Abubucker, S., Badger, J. H., Chinwalla, A. T., Earl, A. M., FitzGerald, M. G., Fulton, R. S., Hallsworth-Pepin, K., Lobos, E. A., Madupu, R., Magrini, V., Martin, J. C., Mitreva, M., Muzny, D. M., Sodergren, E. J., Versalovic, J., Wollam, A. M., Worley, K. C., Wortman, J. R., Young, S. K., Zeng, Q., Aagaard, K. M., Abolude, O. O., Allen-Vercoe, E., Alm, E. J., Alvarado, L., Andersen, G. L.,

Anderson, S., Appelbaum, E., Arachchi, H. M., Armitage, G., Arze, C. A., Ayvaz, T., Baker, C. C., Begg, L., Belachew, T., Bhonagiri, V., Bihan, M., Blaser, M. J., Bloom, T., Bonazzi, V. R., Brooks, P., Buck, G. A., Buhay, C. J., Busam, D. A., Campbell, J. L., Canon, S. R., Cantarel, B. L., Chain, P. S., Chen, I. M., Chen, L., Chhibba, S., Chu, K., Ciulla, D. M., Clemente, J. C., Clifton, S. W., Conlan, S., Crabtree, J., Cutting, M. A., Davidovics, N. J., Davis, C. C., DeSantis, T. Z., Deal, C., Delehaunty, K. D., Dewhirst, F. E., Deych, E., Ding, Y., Dooling, D. J., Dugan, S. P., Dunne, W., Durkin, A., Edgar, R. C., Erlich, R. L., Farmer, C. N., Farrell, R. M., Faust, K., Feldgarden, M., Felix, V. M., Fisher, S., Fodor, A. A., Forney, L., Foster, L., Di Francesco, V., Friedman, J., Friedrich, D. C., Fronick, C. C., Fulton, L. L., Gao, H., Garcia, N., Giannoukos, G., Giblin, C., Giovanni, M. Y., Goldberg, J. M., Goll, J., Gonzalez, A., Griggs, A., Gujja, S., Haas, B. J., Hamilton, H. A., Harris, E. L., Hepburn, T. A., Herter, B., Hoffmann, D. E., Holder, M. E., Howarth, C., Huang, K. H., Huse, S. M., Izard, J., Jansson, J. K., Jiang, H., Jordan, C., Joshi, V., Katancik, J. A., Keitel, W. A., Kelley, S. T., Kells, C., Kinder-Haake, S., King, N. B., Knight, R., Knights, D., Kong, H. H., Koren, O., Koren, S., Kota, K. C., Kovar, C. L., Kyrpides, N. C., La Rosa, P. S., Lee, S. L., Lemon, K. P., Lennon, N., Lewis, C. M., Lewis, L., Ley, R. E., Li, K., Liolios, K., Liu, B., Liu, Y., Lo, C. C., Lozupone, C. A., Lunsford, R., Madden, T., Mahurkar, A. A., Mannon, P. J., Mardis, E. R., Markowitz, V. M., Mavrommatis, K., McCorrison, J. M., McDonald, D., McEwen, J., McGuire, A. L., McInnes, P., Mehta, T., Mihindukulasuriya, K. A., Miller, J. R., Minx, P. J., Newsham, I., Nusbaum, C., O’Laughlin, M., Orvis, J., Pagani, I., Palaniappan, K., Patel, S. M., Pearson, M., Peterson, J., Podar, M., Pohl, C., Pollard, K. S., Priest, M. E., Proctor, L. M., Qin, X., Raes, J., Ravel, J., Reid, J. G., Rho, M., Rhodes, R., Riehle, K. P., Rivera, M. C., Rodriguez-Mueller, B., Rogers, Y. H., Ross, M. C., Russ, C., Sanka, R. K., Sankar, P., Sathirapongsasuti, J., Schloss, J. A., Schloss, P. D., Schmidt, T. M., Scholz, M., Schriml, L., Schubert, A. M., Segata, N., Segre, J. A., Shannon, W. D., Sharp, R. R., Sharpton, T. J., Shenoy, N., Sheth, N. U., Simone, G. A., Singh, I., Smillie, C. S., Sobel, J. D., Sommer, D. D., Spicer, P., Sutton, G. G., Sykes, S. M., Tabbaa, D. G., Thiagarajan, M., Tomlinson, C. M., Torralba, M.,

- Treangen, T. J., Truty, R. M., Vishnivetskaya, T. A., Walker, J., Wang, L., Wang, Z., Ward, D. V., Warren, W., Watson, M. A., Wellington, C., Wetterstrand, K. A., White, J. R., Wilczek-Boney, K., Wu, Y. Q., Wylie, K. M., Wylie, T., Yandava, C., Ye, L., Ye, Y., Yooseph, S., Youmans, B. P., Zhang, L., Zhou, Y., Zhu, Y., Zoloth, L., Zucker, J. D., Birren, B. W., Gibbs, R. A., Highlander, S. K., Weinstock, G. M., Wilson, R. K., and White, O. (2012). A framework for human microbiome research. *Nature*, **486**(7402), 215–221.
- [14] Moyerbrailean, G. A., Davis, G. O., Harvey, C. T., Watzka, D., Wen, X., Pique-Regi, R., and Luca, F. (2015). A high-throughput RNA-seq approach to profile transcriptional responses. *Sci Rep*, **5**, 14976.
- [15] Navas-Molina, J. A., Peralta-Sanchez, J. M., Gonzalez, A., McMurdie, P. J., Vazquez-Baeza, Y., Xu, Z., Ursell, L. K., Lauber, C., Zhou, H., Song, S. J., Huntley, J., Ackermann, G. L., Berg-Lyons, D., Holmes, S., Caporaso, J. G., and Knight, R. (2013). Advancing our understanding of the human microbiome using QIIME. *Meth. Enzymol.*, **531**, 371–444.
- [16] Richards, A. L., Burns, M. B., Alazizi, A., Barreiro, L. B., Pique-Regi, R., Blekhman, R., and Luca, F. (2016). Genetic and transcriptional analysis of human host response to healthy gut microbiota. *mSystems*, **1**(4).
- [17] Turnbaugh, P. J., Backhed, F., Fulton, L., and Gordon, J. I. (2008). Diet-induced obesity is linked to marked but reversible alterations in the mouse distal gut microbiome. *Cell Host Microbe*, **3**(4), 213–223.
- [18] Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., Klemm, A., Flicek, P., Manolio, T., Hindorf, L., and Parkinson, H. (2014). The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.*, **42**(Database issue), D1001–1006.