*Supplementary Material*

— — —

# Gut microbiota has a widespread and modifiable effect on host gene regulation

Allison L Richards [1], Amanda L Muehlbauer [3,4], Adnan Alazizi [1], Michael B Burns [3,4], Anthony Findley [1], Francesco Messina [1], Trevor J Gould [3,4], Camilla Cascardo [1], Roger Pique-Regi [1,2], Ran Blekhman [3,4,5], Francesca Luca [1,2,5]

[1]Center for Molecular Medicine and Genetics, and

[2]Department of Obstetrics and Gynecology, Wayne State University, Detroit, Michigan 48201, USA

[3]Department of Genetics, Cell Biology and Development, and

[4]Department of Ecology, Evolution and Behavior, University of Minnesota, Minneapolis, Minnesota, USA

[5]To whom correspondence should be addressed.

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

October 14, 2018

# Extended Materials and Methods

## 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% CO2. 24 hours before treatment, cells were changed to antibiotic-free media and moved to an incubator at 37°C, 5% $CO_2$, and 5% $O_2$ 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 [1]. 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 [1].

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, foward: 5'-CTTTCAGCAGGGAAGAGTCAA-3', reverse: 5'-AGCCATGCACCACCTGTATGG-3' [2]. The PCR product was run on agarose gel with a 100bp ladder (NEB, cat#: N3231S). We observed the expected band of 590bp (Figure S14) 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'-CGGTTACCTTGTTACGACTT-3' [2].

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 [3]. 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 [4] (`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
```
(1)

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 removed reads with a quality score of $< 10$ (equating to reads mapped to multiple locations) and removed duplicate reads using `samtools rmdup` (`http://github.com/samtools/`).

## Differential Gene Expression Analysis

To identify differentially expressed (DE) genes, we used `DESeq2` [5] (`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:

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

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 $\{1, 2 and 4\}$, $M_n$ 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 $< 10\%$, $|\text{logFC}| > 0.25$ (Figure S4).

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:

$$\begin{aligned} &\text{Gene expression} \sim \text{time point} + \text{microbiota sample} \\ &Y_{jn} = \sum_t \tau_{jt} T_{tn} + \sum_{tm} \beta_{jtm}^M M_{tmn} \end{aligned} \tag{3}$$

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 $< 10\%$, $|\text{logFC}| > 0.25$ (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:

$$\begin{gathered}\text{Gene expression} \sim \text{time point} + \text{treatment} + \\ \text{baseline alpha diversity} \\ Y_{jn} = \sum_t \tau_{jt} T_{tn} + \beta_j^M M_n + \sum_m \alpha_j A_m M_{mn}\end{gathered} \tag{4}$$

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 $g$:

$$\begin{gathered}\text{Gene expression} \sim \text{time point} + \text{treatment} + \\ \text{baseline abundance} \\ Y_{jn} = \sum_t \tau_{jt} T_{tn} + \beta_j^M M_n + \sum_m \gamma_{jg} G_{gm} M_{mn}\end{gathered} \tag{5}$$

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 S7).

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:

$$\begin{aligned} &\text{Gene expression} \sim \text{treatment with microbiome} + \\ &\qquad\qquad\qquad\qquad \text{relative abundance of } \textit{Collinsella aerofaciens} \\ &Y_{jn} = \mu_j + \beta_j^M M_n + \gamma_j^C G_n^C M_n \end{aligned} \tag{6}$$

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 S10, Figure S8) 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 [6] 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 S4, Figure S5).

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

We downloaded the GWAS catalog [7] (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 S8.

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 2A 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 2D 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. [8]. 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**RGGATTAGATACCC**-3', and reverse primer 5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG**CGACRRCCATGCANCACCT**-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'-CAAGCAGAAGACGGCATACGAGAT[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 $>$ 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 [9,

10]. OTUs were picked using the closed reference algorithm against the Greengenes database (August, 2013 release) [8, 9, 10, 11]. 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 S1B).

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 [12, 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 S2). 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

$$\begin{gathered}\text{Transformed Relative Abundance of Taxa} \sim \text{Individual} + \text{Time} \\ G_{gn} = \sum_t \tau_{gt} T_{tn} + \sum_m \beta^M_{gm} M_{mn}\end{gathered} \tag{7}$$

Alternative Model

$$\begin{gathered}\text{Transformed Relative Abundance of Taxa} \sim \text{Individual} + \\ \text{Treatment} + \text{Time} \\ G_{gn} = \sum_t \tau_{gt} T_{tn} + \sum_m \beta^M_{gm} M_{mn} + \gamma_g K_n\end{gathered} \tag{8}$$

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 $\beta^M_{gm}$ 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

16S rRNA sequencing data and RNA sequencing data of colonocytes in all conditions have been deposited to the SRA (SRP150967).

# Additional Results and Analyses

## 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 1.8.0 was used to analyze the resulting sequences [9]. 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 S1A and C). The microbiome maintains its composition during culture and our sequencing depth is able to capture this variation (Figure S1B). 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 $< 10\%$, $|\log_2 \mathrm{FC}| > 0.25$) 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 S4A and B, Table S3). 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 S6, 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 [14, 15, 16, 17, 18].

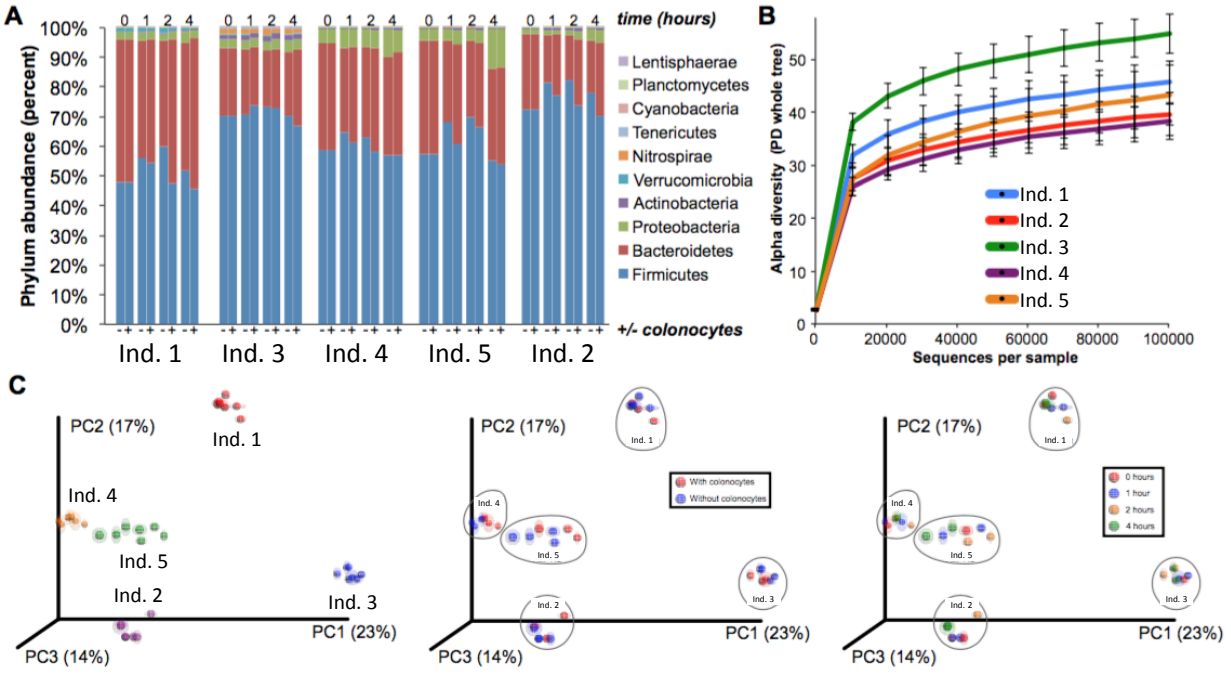


Figure S1: **Analysis of the Microbial Community Changes Over Time When Co-cultured with Normal Human Colonocytes. A** Relative abundances of the bacterial phyla in each of the five different microbial communities at 0 (uncultured), 1, 2 and 4 hours with and without co-culturing (+/-) with colonocytes. **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 unifrac) 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.

Table S1: **16S Sequencing Counts Reduced to Genera-Level.** Counts generated from 16S sequencing of all samples, including the uncultured samples. The taxonomy is reduced to the genus level and was used for the genera-based model to determine association of host gene expression and abundance of various genera.

(see TableS1.txt)

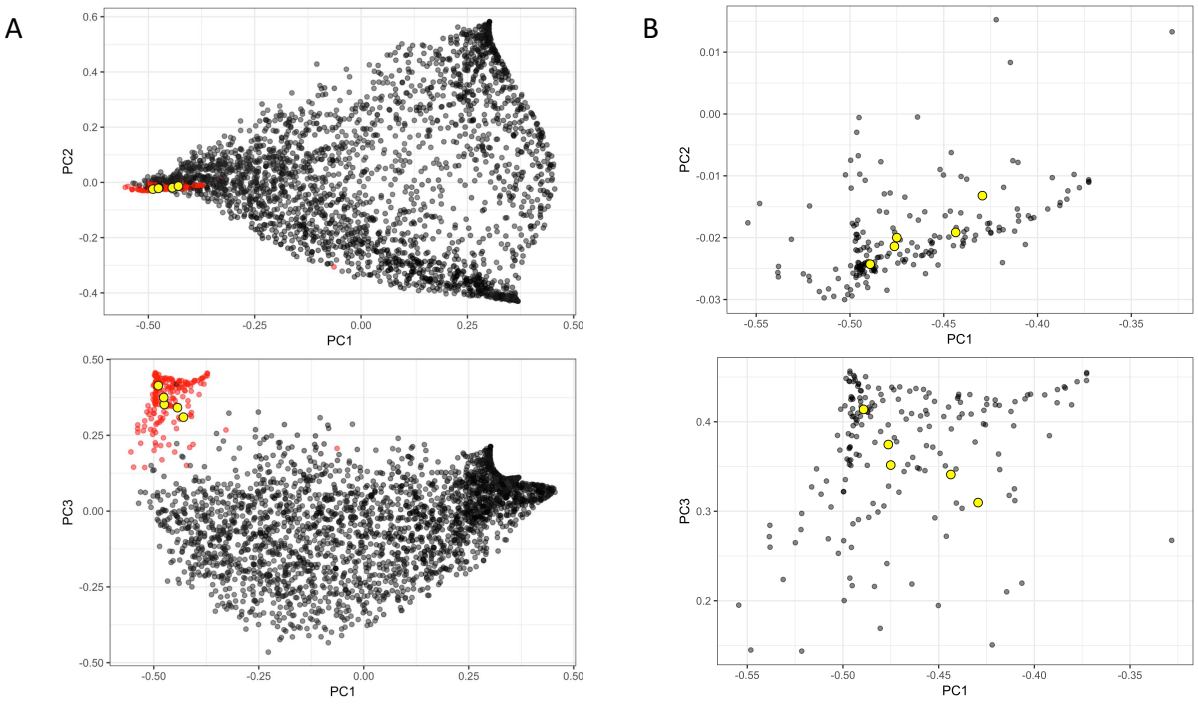


Figure S2: **Comparison of Fecal Samples to Human Microbiome Project. A** PCoA representing samples from HMP and the current study, with HMP fecal samples in red and all other HMP samples in grey. The five fecal samples used in the current study are colored yellow. Top panel shows PC1 (x-axis) vs PC2 (y-axis), while bottom panel shows PC1 (x-axis) vs PC3 (y-axis). **B** Same plots focusing on only the HMP fecal samples (now in gray) and the samples in the current study.

Table S2: **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.
(see TableS2.txt)

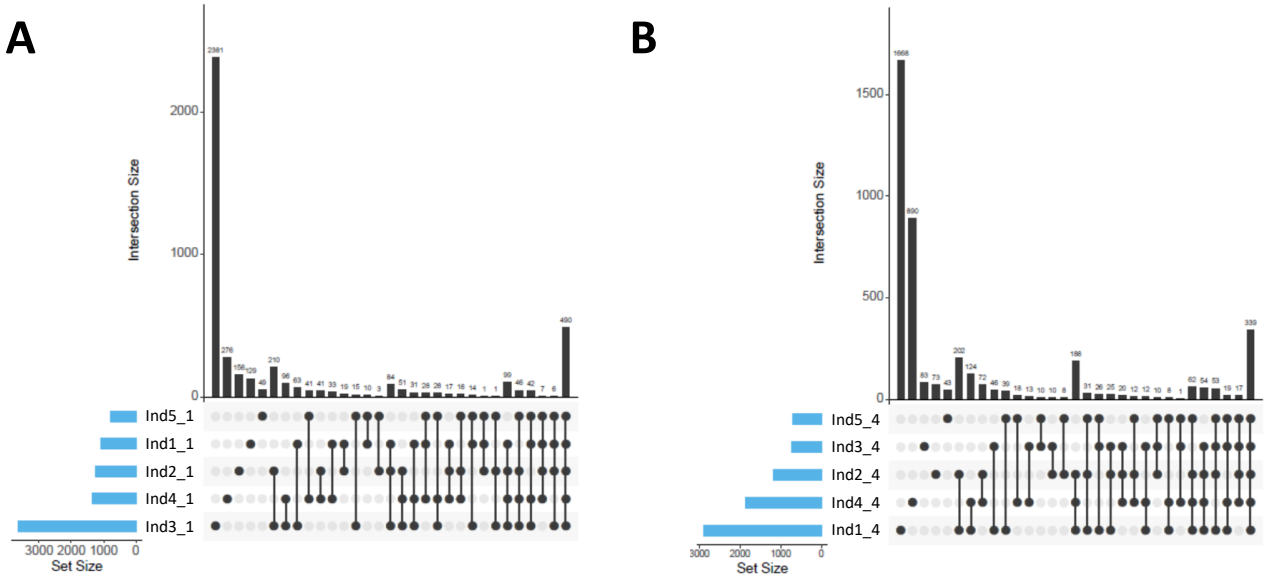


Figure S3: **Comparison of Differentially Expressed Genes Across Time Points and Microbiota.** **A** These plots were generated by UpSetR package to show the comparison of genes that are differentially expressed at 1 hour following exposure with each of the five microbiota. **B** shows the comparison of differentially expressed genes across the five treatments at 4hr.

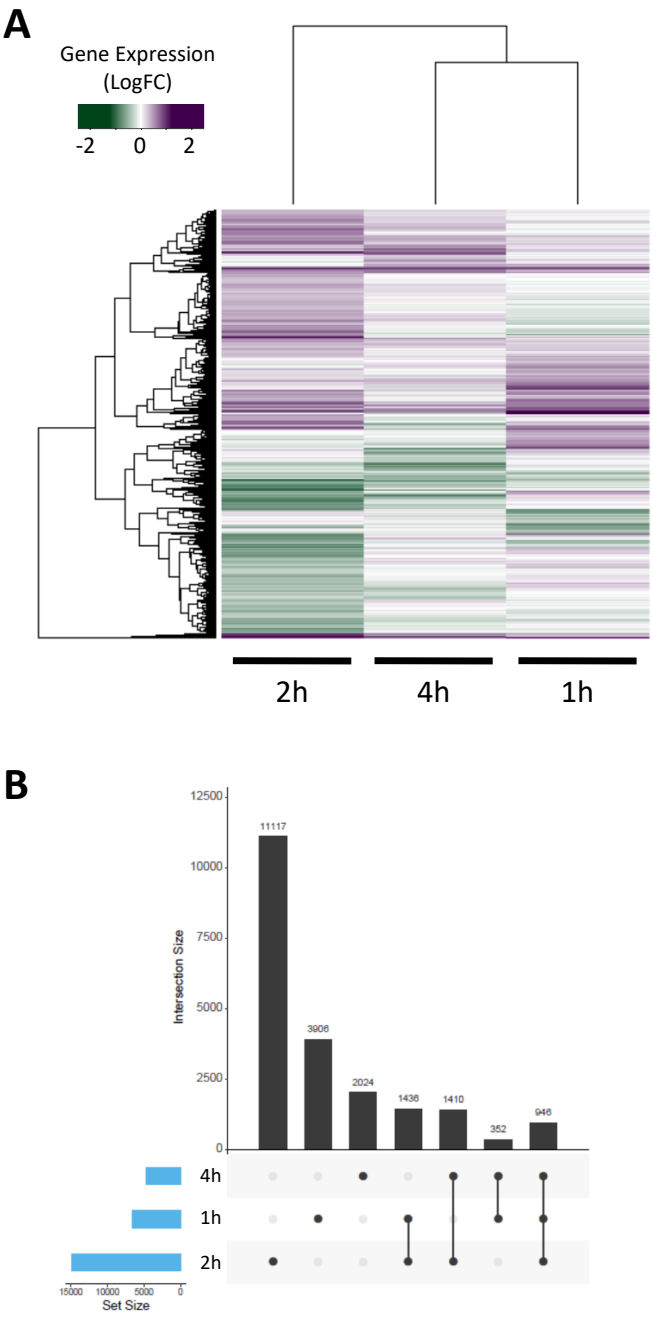


Figure S4: **Differentially Expressed Genes Across Time Points.** **A** This heatmap shows gene expression changes from the model in which the five microbiota treatments at a given time point are used as replicates. Purple denotes an increase in gene expression (green shows a reduction) compared to the gene expression in colonocytes cultured alone. Only genes that are differentially expressed in at least one time point 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.

Table S3: **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.

(see TableS3.txt)

Table S4: **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 | 0.00057533 |
| cell projection organization | 120.908 | 160 | 0.00057533 |
| 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 |

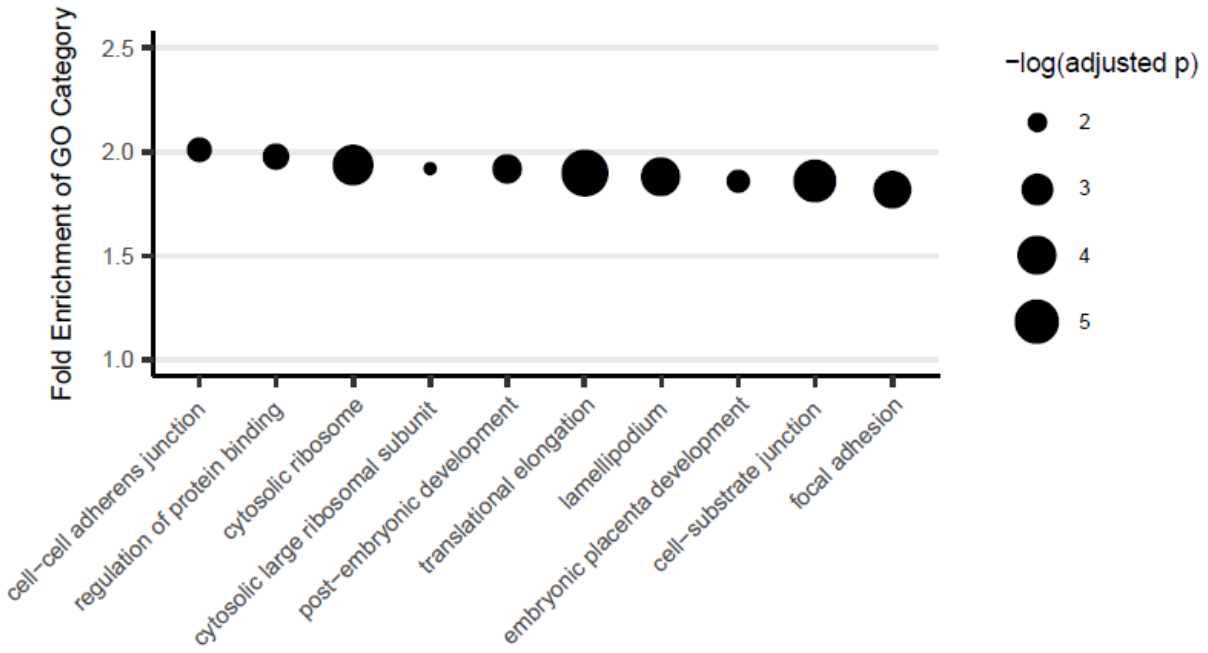


Figure S5: **Gene Ontology for DEG at any Time Point** Plot showing gene ontology enrichment for genes that were differentially expressed at any time point in the model in which each treatment is considered a replicate (and treatments lead to similar expression changes). The top 10 enrichments are shown in this plot, excluding categories where the expected number of genes is less than 10 or greater than 500. The size of the point is the $-\log_{10}$ of the adjusted $p$-value.

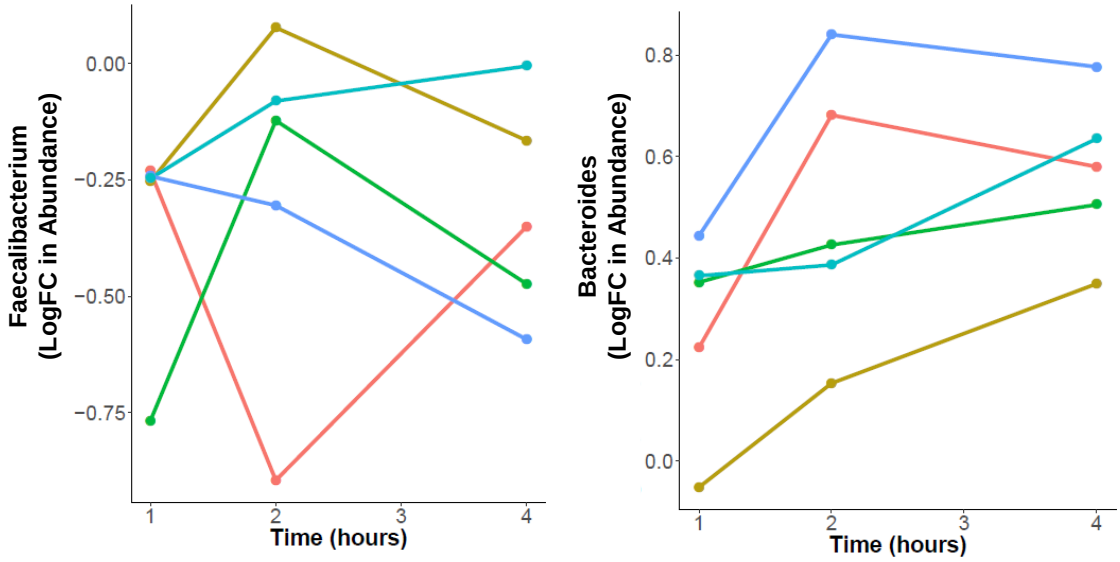


Figure S6: **Examples of Taxa Affected by Culturing with Colonocytes.** Genera (*Faecalibacterium* and *Bacteroides*) whose changes in abundance following co-culturing is inconsistent dependent on the microbiota sample in which it is found. Each color denotes samples exposed to a different microbiota sample: Ind1 (red), Ind2 (blue), Ind3 (yellow), Ind4 (green) and Ind5 (teal).

| Bacteria Taxa | p-value | BH adjusted p-value |
|---|---|---|
| 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: **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.

(see TableS6.txt)

Table S7: **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.

(see TableS7.txt)

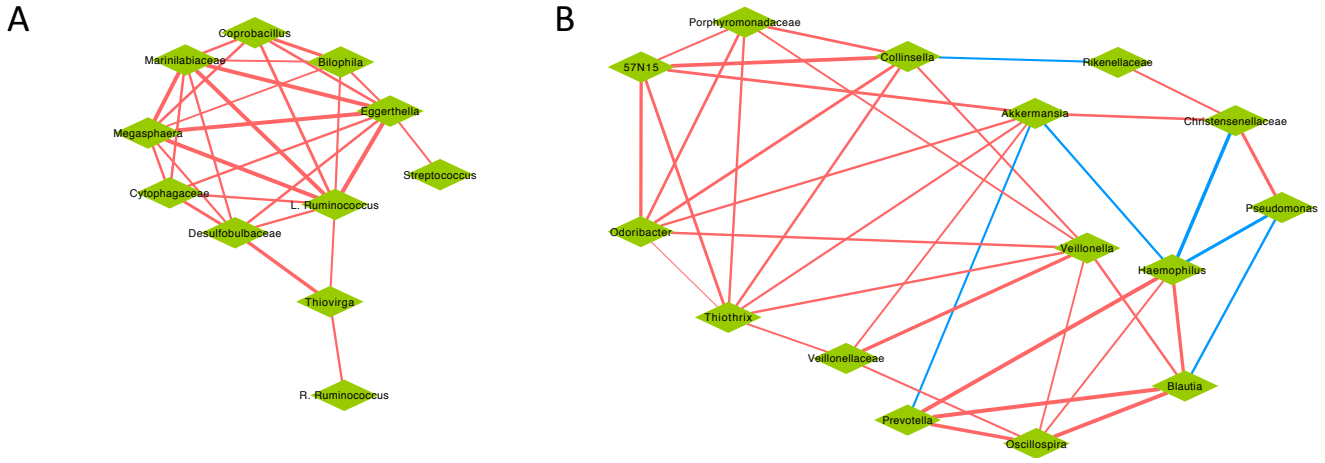


Figure S7: **Network of Microbes.** Correlations within microbial taxa clusters shown in Figure 2A. **A** shows microbes positively associated with genes in cluster 1, while **B** shows microbes associated with genes in cluster 2. Red edges indicate a positive correlation. Blue edges indicate a negative correlation. Edge thickness indicates the magnitude of the correlation. Correlations and corresponding significance between pairs of microbial taxa were generated using SparCC, which estimates linear Pearson correlations, with 100 bootstrap iterations [19]. Correlations were filtered by BH FDR $< 5\%$ and by correlation magnitude $> 0.5$.

| 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 S8: **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 S9: **DE Genes from Taxon Model Found Associated with GWAS Traitsl.** Table shows gene symbol and Ensembl ID for genes identified and differentially expressed in the genera-based model. The third column shows the genera with which gene expression was associated and the fourth column delineates the traits and papers where the gene was associated with a GWAS trait.

| Gene Symbol | ENSG | Genera | GWAS |
|---|---|---|---|
| PRKCH | ENSG00000027075 | Odoribacter, Streptococcus, Thiothrix | Obesity (early onset extreme) (PMID:23563609), Reading and spelling (PMID:23738518), Rheumatoid arthritis (PMID:22446963), Rheumatoid arthritis (PMID:24390342), Word reading (PMID:23738518) |
| ALG1 | ENSG00000033011 | Streptococcus | Phospholipid levels (plasma) (PMID:22359512) |
| CDH1 | ENSG00000039068 | Collinsella | Colorectal cancer (PMID:19011631), Ulcerative colitis (PMID:19915572) |
| LAMA3 | ENSG00000053747 | Collinsella, Streptococcus, Thiovirga | Allergic rhinitis (PMID:25085501), Amyotrophic lateral sclerosis (PMID:22959728) |
| NEBL | ENSG00000078114 | Odoribacter, Streptococcus, Thiothrix | Ovarian cancer (PMID:23535730) |
| TMED2 | ENSG00000086598 | Collinsella, Thiovirga | Educational attainment (PMID:25201988) |
| ZFAND6 | ENSG00000086666 | Thiovirga | Type 2 diabetes (PMID:20581827) |
| ASAP3 | ENSG00000088280 | Odoribacter, Thiothrix | Cholesterol, total (PMID:24097068) |
| ABLIM1 | ENSG00000099204 | Streptococcus, Thiovirga | Obesity-related traits (PMID:23251661), Prostate cancer (gene x gene interaction) (PMID:22219177) |
| SLC7A5 | ENSG00000103257 | Thiovirga | Blood metabolite levels (PMID:24816252) |
| PIEZO1 | ENSG00000103335 | Collinsella | Red blood cell traits (PMID:23222517) |
| RHOV | ENSG00000104140 | Thiovirga | Visceral adipose tissue/subcutaneous adipose tissue ratio (PMID:22589738) |
| PLAT | ENSG00000104368 | Odoribacter, Streptococcus, Thiothrix | Plasma plasminogen activator levels (PMID:24578379) |
| TYRP1 | ENSG00000107165 | Thiovirga | Amyotrophic lateral sclerosis (sporadic) (PMID:24529757), Blue vs. green eyes (PMID:18488028), Eye color (PMID:23548203), Hair color (PMID:22556244) |
| CUEDC2 | ENSG00000107874 | Thiovirga | Autism spectrum disorder-related traits (PMID:25534755), Prostate cancer (PMID:23535732) |
| MAP3K8 | ENSG00000107968 | Odoribacter, Streptococcus, Thiothrix, Thiovirga | Inflammatory bowel disease (PMID:23128233) |
| DKK1 | ENSG00000107984 | Odoribacter, Thiothrix | Bone properties (heel) (PMID:24430505), Electrocardiographic traits (PMID:20062063), Periodontitis (Mean PAL) (PMID:24024966), QRS duration (PMID:21076409) |
| KRT23 | ENSG00000108244 | Collinsella, Streptococcus, Thiovirga | Height (PMID:19893584) |
| COL1A1 | ENSG00000108821 | Collinsella | Breast cancer (PMID:17903305), Response to taxane treatment (docetaxel) (PMID:23006423) |

| Gene Symbol | ENSG | Genera | GWAS |
|---|---|---|---|
| KRR1 | ENSG00000111615 | Thiothrix | Blood trace element (Cu levels) (PMID:23720494) |
| RPS12 | ENSG00000112306 | Blautia | Type 2 diabetes nephropathy (PMID:21150874) |
| VEGFA | ENSG00000112715 | Odoribacter, Thiothrix | Adiponectin levels (PMID:22479202), Age-related macular degeneration (PMID:21665990), Age-related macular degeneration (PMID:23455636), Chronic kidney disease (PMID:20383146), Coronary heart disease (PMID:22319020), HDL cholesterol (PMID:24097068), Sexual dimorphism in anthropometric traits (PMID:23754948), Thyroid hormone levels (PMID:23408906), Triglycerides (PMID:24097068), Type 2 diabetes (PMID:18372903), Type 2 diabetes (PMID:24509480), Urate levels (PMID:23263486), Waist-hip ratio (PMID:20935629) |
| FN1 | ENSG00000115414 | Collinsella, Streptococcus, Thiovirga | LDL cholesterol (PMID:24097068) |
| LPHN2 | ENSG00000117114 | Odoribacter, Thiothrix | Metabolite levels (X-11787) (PMID:23934736), Obesity-related traits (PMID:23251661), Spontaneous preterm birth (preterm birth) (PMID:25599974), Temperament (PMID:22832960), Toenail selenium levels (PMID:25343990) |
| PRRC2C | ENSG00000117523 | Collinsella | Alzheimer's disease (cognitive decline) (PMID:23535033) |
| F3 | ENSG00000117525 | Odoribacter, Thiothrix | D-dimer levels (PMID:21502573), End-stage coagulation (PMID:23381943), Type 2 diabetes (PMID:22238593) |
| DENND1A | ENSG00000119522 | Odoribacter, Thiothrix | Adverse response to chemotherapy in breast cancer (alopecia) (cyclophosphamide+doxorubicin+/-5FU) (PMID:24025145), Body mass index in non-asthmatics (PMID:23517042), Myopia (pathological) (PMID:23049088), Personality dimensions (PMID:20691247), Polycystic ovary syndrome (PMID:21151128), Waist circumference (PMID:23966867) |
| VPS4B | ENSG00000119541 | Thiovirga | Pyoderma gangrenosum in inflammatory bowel disease (PMID:24487271) |
| CSTA | ENSG00000121552 | Thiovirga | Calcium levels (PMID:20705733), Calcium levels (PMID:25972035) |
| DSTN | ENSG00000125868 | Blautia | Presence of antiphospholipid antibodies (PMID:23509613) |
| CANX | ENSG00000127022 | Collinsella | Resting heart rate (PMID:24680774) |
| ARL8B | ENSG00000134108 | Thiovirga | Disc degeneration (lumbar) (PMID:22993228) |

| Gene Symbol | ENSG | Genera | GWAS |
|---|---|---|---|
| FADS2 | ENSG00000134824 | Blautia | Cholesterol, total (PMID:19060911), Cholesterol, total (PMID:20686565), Cholesterol, total (PMID:24097068), Colorectal cancer (PMID:24836286), Comprehensive strength and appendicular lean mass (PMID:22960237), Glycated hemoglobin levels (PMID:24647736), HDL cholesterol (PMID:19060906), HDL cholesterol (PMID:19060911), HDL cholesterol (PMID:20686565), HDL cholesterol (PMID:24097068), Inflammatory bowel disease (PMID:23128233), Iron status biomarkers (transferrin levels) (PMID:25352340), Laryngeal squamous cell carcinoma (PMID:25194280), LDL cholesterol (PMID:19060910), LDL cholesterol (PMID:19060911), LDL cholesterol (PMID:20686565), LDL cholesterol (PMID:24097068), Lipid metabolism phenotypes (PMID:22286219), Liver enzyme levels (alkaline phosphatase) (PMID:22001757), Metabolic syndrome (PMID:20694148), Metabolite levels (PMID:22916037), Metabolite levels (PMID:23378610), Oleic acid (18:1n-9) plasma levels (PMID:23362303), Palmitoleic acid (16:1n-7) plasma levels (PMID:23362303), Phospholipid levels (plasma) (PMID:21829377), Phospholipid levels (plasma) (PMID:22359512), Plasma omega-6 polyunsaturated fatty acid levels (adrenic acid) (PMID:24823311), Plasma omega-6 polyunsaturated fatty acid levels (arachidonic acid) (PMID:24823311), Plasma omega-6 polyunsaturated fatty acid levels (dihomo-gamma-linolenic acid) (PMID:24823311), Plasma omega-6 polyunsaturated fatty acid levels (gamma-linolenic acid) (PMID:24823311), Plasma omega-6 polyunsaturated fatty acid levels (linoleic acid) (PMID:24823311), P wave duration (PMID:24850809), QT interval (PMID:24952745), Red blood cell fatty acid levels (PMID:25500335), Response to statin therapy (PMID:20339536), Rheumatoid arthritis (PMID:24390342), Sphingolipid levels (PMID:22359512), Stearic acid (18:0) plasma levels (PMID:23362303), Triglycerides (PMID:19060906), Triglycerides (PMID:20686565), Triglycerides (PMID:24097068) |

| Gene Symbol | ENSG | Genera | GWAS |
|---|---|---|---|
| ANXA1 | ENSG00000135046 | Blautia | Schizophrenia, bipolar disorder and depression (combined) (PMID:20713499) |
| LAMC1 | ENSG00000135862 | Collinsella | Clozapine-induced agranulocytosis (PMID:25187353), Colorectal cancer (PMID:23266556), Systemic lupus erythematosus (PMID:24871463) |
| MYC | ENSG00000136997 | Odoribacter, Thiothrix | Allergic sensitization (PMID:23817571), Bladder cancer (PMID:20972438), Breast cancer (PMID:23535729), Chronic lymphocytic leukemia (PMID:23770605), Colorectal cancer (PMID:23266556), Colorectal cancer (PMID:24737748), Diffuse large B cell lymphoma (PMID:25261932), Height (PMID:25429064), Multiple sclerosis (PMID:21833088), Ovarian cancer (PMID:20852632), Ovarian cancer (PMID:23535730), Pancreatic cancer (PMID:25086665), Prostate cancer (early onset) (PMID:24740154), Prostate cancer (PMID:24753544), Renal cell carcinoma (PMID:24220699), Urinary bladder cancer (PMID:18794855), Urinary bladder cancer (PMID:20348956), Urinary bladder cancer (PMID:24861552) |
| FDX1 | ENSG00000137714 | Odoribacter, Thiothrix | Cognitive function (PMID:24684796) |
| SLC5A6 | ENSG00000138074 | Odoribacter, Streptococcus, Thiothrix, Thiovirga | Blood metabolite levels (PMID:24816252) |
| GLTP | ENSG00000139433 | Blautia, Collinsella, Thiovirga | Metabolic syndrome (PMID:20694148) |
| GAREM | ENSG00000141441 | Streptococcus | PR interval (PMID:25035420) |
| MAL2 | ENSG00000147676 | Collinsella, Streptococcus, Thiovirga | Blood pressure (smoking interaction) (PMID:25189868), Hippocampal atrophy (PMID:19668339), Osteoprotegerin levels (PMID:25080503) |
| MKI67 | ENSG00000148773 | Collinsella | Select biomarker traits (PMID:17903293) |
| TMEM219 | ENSG00000149932 | Collinsella | Schizophrenia (PMID:25056061) |
| ITGB1 | ENSG00000150093 | Thiovirga | Depression (quantitative trait) (PMID:20800221), Suicide in bipolar disorder (PMID:25917933) |
| CAST | ENSG00000153113 | Thiovirga | Alcohol dependence (PMID:19581569) |
| S100P | ENSG00000163993 | Thiovirga | Obesity-related traits (PMID:23251661) |
| RPS14 | ENSG00000164587 | Blautia | Visceral adipose tissue/subcutaneous adipose tissue ratio (PMID:22589738) |
| RPL27A | ENSG00000166441 | Blautia | Body mass index (PMID:20935630), Obesity (PMID:23563607) |
| A2ML1 | ENSG00000166535 | Odoribacter, Thiothrix | Cholesterol, total (PMID:24097068) |
| HSP90B1 | ENSG00000166598 | Streptococcus | Coronary heart disease (PMID:21626137) |
| ANPEP | ENSG00000166825 | Thiovirga | Blood metabolite levels (PMID:24816252) |

| Gene Symbol | ENSG | Genera | GWAS |
|---|---|---|---|
| UGT1A6 | ENSG00000167165 | Thiothrix | Total bilirubin levels in HIV-1 infection (PMID:25884002) |
| MMADHC | ENSG00000168288 | Blautia | Liver enzyme levels (alanine transaminase) (PMID:24124411) |
| IL7R | ENSG00000168685 | Collinsella, Streptococcus | Multiple sclerosis (PMID:19525953), Multiple sclerosis (PMID:21244703), Multiple sclerosis (PMID:21833088), Primary biliary cirrhosis (PMID:21399635), Primary biliary cirrhosis (PMID:23000144), Type 1 diabetes (PMID:17554260), Ulcerative colitis (PMID:21297633) |
| PCDH7 | ENSG00000169851 | Odoribacter | Electroencephalogram traits (PMID:25387704), Epilepsy (PMID:25087078), Lipid traits (PMID:22028671), Menarche (age at onset) (PMID:25231870), Obesity-related traits (PMID:23251661), Refractive astigmatism (PMID:25367360), Underweight status (PMID:25133637), Very long-chain saturated fatty acid levels (fatty acid 22:0) (PMID:25378659) |
| RPL15 | ENSG00000174748 | Blautia | Acne (severe) (PMID:24927181) |
| AIM1L | ENSG00000176092 | Streptococcus, Thiovirga | Obesity-related traits (PMID:23251661) |
| NFATC2IP | ENSG00000176953 | Odoribacter | Educational attainment (PMID:25201988) |
| HIGD1A | ENSG00000181061 | Thiovirga | Clozapine-induced agranulocytosis (PMID:25187353), Obesity-related traits (PMID:23251661) |
| SORCS2 | ENSG00000184985 | Odoribacter, Thiothrix | Coronary artery calcification (PMID:23870195), Hippocampal sclerosis (PMID:25188341), Insulin-like growth factors (PMID:21216879), Obesity-related traits (PMID:23251661) |
| NAP1L1 | ENSG00000187109 | Blautia | Obesity-related traits (PMID:23251661) |
| MT1X | ENSG00000187193 | Blautia | Gambling (PMID:22780124) |
| DNM3 | ENSG00000197959 | Odoribacter, Thiothrix | Bone mineral density (PMID:22504420), Height (PMID:18391951), Height (PMID:20881960), Height (PMID:23563607), Height (PMID:25282103), Height (PMID:25429064), Mean platelet volume (PMID:19820697), Mean platelet volume (PMID:22139419), Mean platelet volume (PMID:24026423), Platelet counts (PMID:22139419), Very long-chain saturated fatty acid levels (fatty acid 22:0) (PMID:25378659), Waist-hip ratio (PMID:20935629) |
| UBL5 | ENSG00000198258 | Blautia | Thiazide-induced adverse metabolic effects in hypertensive patients (PMID:23400010) |

| Gene Symbol | ENSG | Genera | GWAS |
|---|---|---|---|
| PRRC2A | ENSG00000204469 | Thiovirga | Blood metabolite ratios (PMID:24816252), Menopause (age at onset) (PMID:22267201), Schizophrenia (PMID:23894747) |
| DENND1B | ENSG00000213047 | Odoribacter | Asthma (PMID:20032318), Crohn's disease (PMID:21102463), Primary biliary cirrhosis (PMID:21399635), Serum protein levels (sST2) (PMID:23999434) |
| MIF | ENSG00000240972 | Blautia | Liver enzyme levels (gamma-glutamyl transferase) (PMID:22001757) |

Table S10: **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.

(see TableS10.txt)

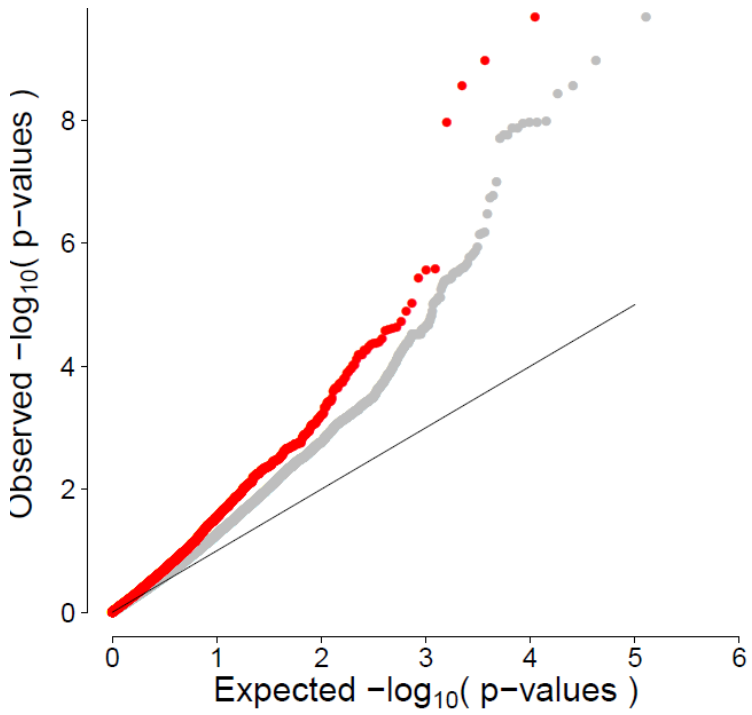


Figure S8: **Comparison of Taxon Model with *Collinsella* and Spike-In of *Collinsella aerofaciens*.** QQ-plot of $p$-values from the model of the five microbiota experiment with abundance of baseline *Collinsella* (gray). The red points indicate the same values from the gray but subsetted to only include 1,570 genes that are DEG in the validation spike-in experiment with *Collinsella aerofaciens*.

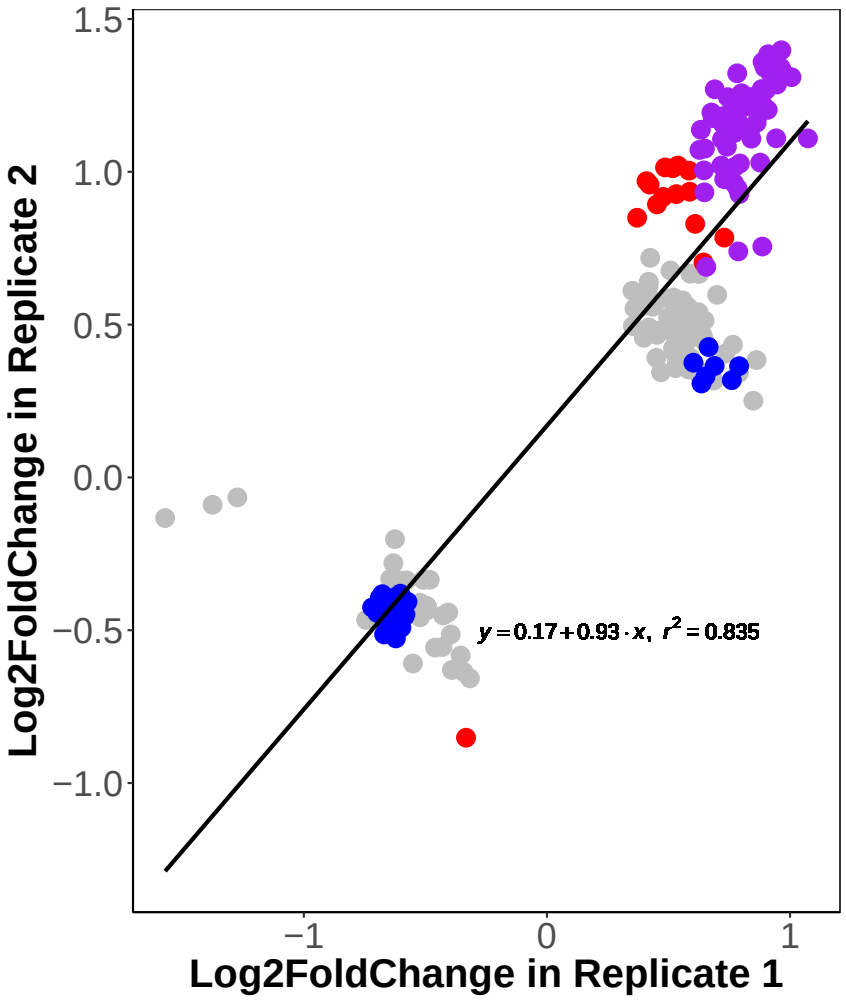


Figure S9: **Comparison of Differential Chromatin Accessibility Across Replicates.** We identified changes in chromatin accessibility between microbiome-exposed samples and control samples for each replicate and show that these changes are highly correlated ($r^2 = 0.835$, $p$-value $< 2.2 \times 10^{-16}$). Points here denote positions that we identified as being differentially accessible when the replicates are used together in the same model. Gray points show regions that are not considered differentially accessible in either replicate when considered alone. Blue and red points are regions that are differentially accessible in one replicate (BH FDR $< 20\%$) and purple points are regions that even when the replicates are considered separately, are considered differentially accessible in both.

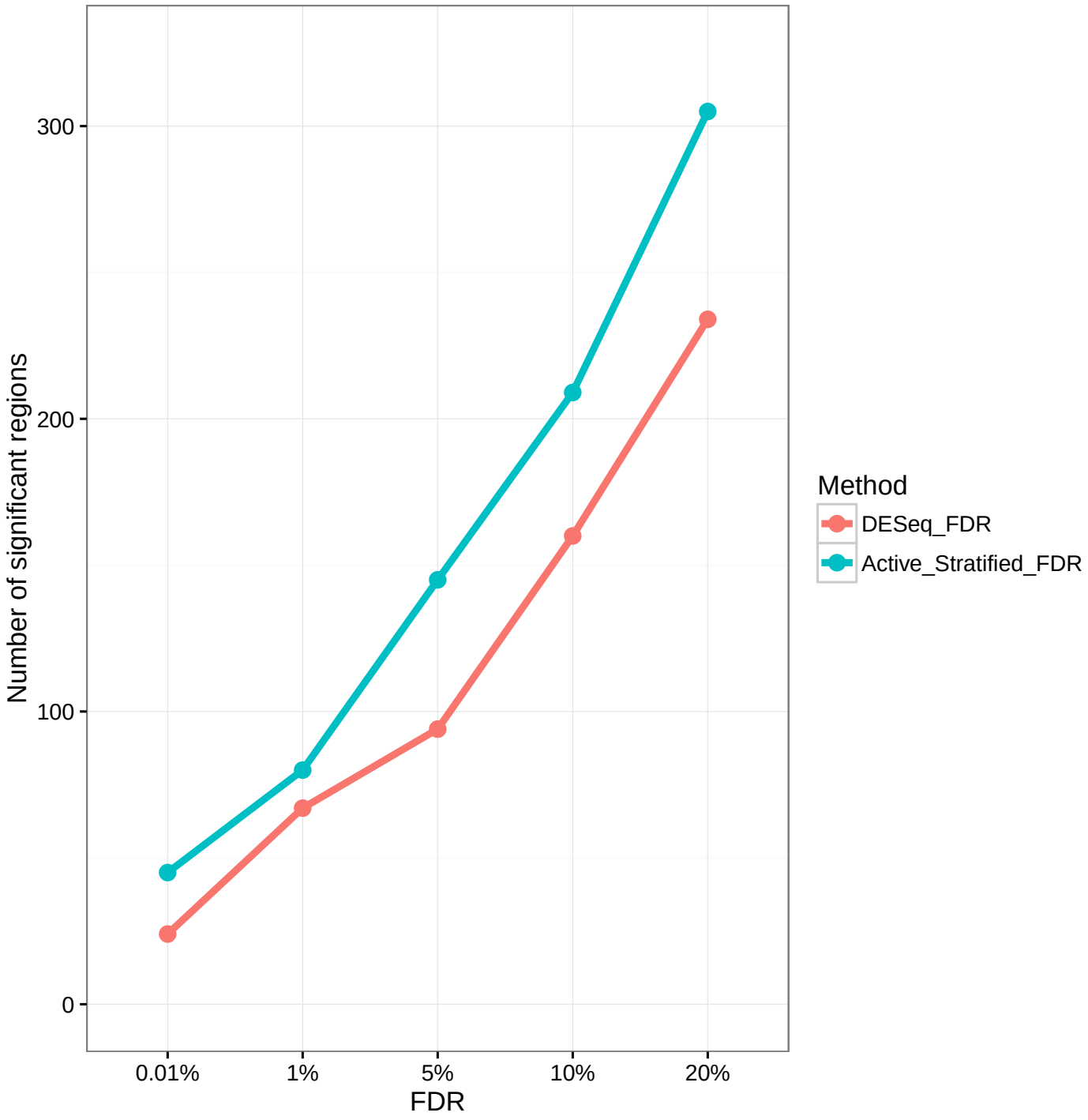


Figure S10: **Comparing Thresholds and Models for Identifying Differential Chromatin Accessibility.** Plotted are the number of differentially accessible regions at varying thresholds using the DESeq2 model and the stratified FDR approach considering only active motifs.

Table S11: **Differentially Accessible Regions within 50kb of Differentially Expressed Genes.** Table denotes all transcript-by-ATAC region pair for which both region and transcript are differentially expressed following treatment with the microbiome using the model where we consider the 5 microbiota treatments to be replicates.

| Transcript | Gene Symbol | Chr | Transcript TSS | BH p-value for Transcript (4hours) | ATAC Region | BH p-value for ATAC (2hours) |
|---|---|---|---|---|---|---|
| ENST00000473257 | ACTB | chr7 | 5570235 | 0.042541618 | 7:5566500-5566800 | 0.074209361 |
| ENST00000480301 | ACTB | chr7 | 5570214 | 0.050050361 | 7:5566500-5566800 | 0.074209361 |
| ENST00000417101 | ACTB | chr7 | 5569613 | 0.042979454 | 7:5566500-5566800 | 0.074209361 |
| ENST00000425660 | ACTB | chr7 | 5570232 | 0.029293417 | 7:5566500-5566800 | 0.074209361 |
| ENST00000443528 | ACTB | chr7 | 5602459 | 0.069589647 | 7:5566500-5566800 | 0.074209361 |
| ENST00000484841 | ACTB | chr7 | 5570233 | 0.047398859 | 7:5566500-5566800 | 0.074209361 |
| ENST00000462494 | ACTB | chr7 | 5570232 | 0.029293417 | 7:5566500-5566800 | 0.074209361 |
| ENST00000477812 | ACTB | chr7 | 5570221 | 0.045826835 | 7:5566500-5566800 | 0.074209361 |
| ENST00000493945 | ACTB | chr7 | 5569294 | 0.020700658 | 7:5566500-5566800 | 0.074209361 |
| ENST00000432588 | ACTB | chr7 | 5603415 | 0.051756601 | 7:5566500-5566800 | 0.074209361 |
| ENST00000414620 | ACTB | chr7 | 5602421 | 0.092630026 | 7:5566500-5566800 | 0.074209361 |
| ENST00000464611 | ACTB | chr7 | 5567729 | 0.046310197 | 7:5566500-5566800 | 0.074209361 |
| ENST00000331789 | ACTB | chr7 | 5570340 | 0.02926373 | 7:5566500-5566800 | 0.074209361 |
| ENST00000537102 | ANXA11 | chr10 | 81965314 | 1.30E-05 | 10:81956100-81956400 | 0.095911217 |

| | | | | | | |
|---|---|---|---|---|---|---|
| ENST00000535999 | ANXA11 | chr10 | 81964247 | 1.46E-05 | 10:81956100-81956400 | -0.095911217 |
| ENST00000265447 | ANXA11 | chr10 | 81965328 | 5.27E-06 | 10:81956100-81956400 | -0.095911217 |
| ENST00000474545 | ANXA11 | chr10 | 81964675 | 0.005818049 | 10:81956100-81956400 | -0.095911217 |
| ENST00000447489 | ANXA11 | chr10 | 81921768 | 0.004051488 | 10:81956100-81956400 | -0.095911217 |
| ENST00000437799 | ANXA11 | chr10 | 81932777 | 0.02121747 | 10:81956100-81956400 | -0.095911217 |
| ENST00000422982 | ANXA11 | chr10 | 81965328 | 5.27E-06 | 10:81956100-81956400 | -0.095911217 |
| ENST00000360615 | ANXA11 | chr10 | 81965328 | 5.27E-06 | 10:81956100-81956400 | -0.095911217 |
| ENST00000463657 | ANXA11 | chr10 | 81965278 | 0.001151767 | 10:81956100-81956400 | -0.095911217 |
| ENST00000438331 | ANXA11 | chr10 | 81965328 | 3.18E-09 | 10:81956100-81956400 | -0.095911217 |
| ENST00000463340 | ANXA11 | chr10 | 81917774 | 0.011120126 | 10:81956100-81956400 | -0.095911217 |
| ENST00000372231 | ANXA11 | chr10 | 81965328 | 3.18E-09 | 10:81956100-81956400 | -0.095911217 |
| ENST00000535402 | B4GALNT3 | chr12 | 661240 | 0.052645827 | 12:646800-647100 | 0.057233417 |
| ENST00000602675 | C16orf74 | chr16 | 85784557 | 0.001078831 | 16:85783500-85783800 | -1.49E-06 |
| ENST00000602758 | C16orf74 | chr16 | 85784616 | 8.36E-05 | 16:85783500-85783800 | -1.49E-06 |
| ENST00000284245 | C16orf74 | chr16 | 85784697 | 0.001025367 | 16:85783500-85783800 | -1.49E-06 |
| ENST00000602766 | C16orf74 | chr16 | 85784718 | 0.00110132 | 16:85783500-85783800 | -1.49E-06 |
| ENST00000602719 | C16orf74 | chr16 | 85784735 | 8.05E-05 | 16:85783500-85783800 | -1.49E-06 |
| ENST00000602914 | C16orf74 | chr16 | 85784711 | 0.001004739 | 16:85783500-85783800 | -1.49E-06 |
| ENST00000602719 | C16orf74 | chr16 | 85784735 | 8.05E-05 | 16:85785900-85786200 | -0.085944368 |
| ENST00000284245 | C16orf74 | chr16 | 85784697 | 0.001025367 | 16:85785900-85786200 | -0.085944368 |

| | | | | | | |
|---|---|---|---|---|---|---|
| ENST00000602766 | C16orf74 | chr16 | 85784718 | 0.00110132 | 16:85785900-85786200 | -0.085944368 |
| ENST00000602675 | C16orf74 | chr16 | 85784557 | 0.001078831 | 16:85785900-85786200 | -0.085944368 |
| ENST00000602758 | C16orf74 | chr16 | 85784616 | 8.36E-05 | 16:85785900-85786200 | -0.085944368 |
| ENST00000602914 | C16orf74 | chr16 | 85784711 | 0.001004739 | 16:85785900-85786200 | -0.085944368 |
| ENST00000556756 | CIDEB | chr14 | 24775667 | 0.000207256 | 14:24728400-24728700 | -0.189913141 |
| ENST00000554411 | CIDEB | chr14 | 24777469 | 0.000126096 | 14:24728400-24728700 | -0.189913141 |
| ENST00000561281 | DHRS1 | chr14 | 24761127 | 0.049481235 | 14:24728400-24728700 | -0.189913141 |
| ENST00000288111 | DHRS1 | chr14 | 24769039 | 0.061643947 | 14:24728400-24728700 | -0.189913141 |
| ENST00000558340 | DHRS1 | chr14 | 24768949 | 0.064192799 | 14:24728400-24728700 | -0.189913141 |
| ENST00000561273 | DHRS1 | chr14 | 24768967 | 0.066638799 | 14:24728400-24728700 | -0.189913141 |
| ENST00000561137 | DHRS1 | chr14 | 24763862 | 0.056990141 | 14:24728400-24728700 | -0.189913141 |
| ENST00000394773 | EML3 | chr11 | 62380237 | 0.080155431 | 11:62341200-62341500 | -0.060248619 |
| ENST00000453700 | FBXL18 | chr7 | 5553423 | 2.04E-11 | 7:5566500-5566800 | 0.074209361 |
| ENST00000458142 | FBXL18 | chr7 | 5541550 | 1.69E-07 | 7:5566500-5566800 | 0.074209361 |
| ENST00000415009 | FBXL18 | chr7 | 5553383 | 3.54E-25 | 7:5566500-5566800 | 0.074209361 |
| ENST00000382368 | FBXL18 | chr7 | 5553429 | 8.52E-18 | 7:5566500-5566800 | 0.074209361 |
| ENST00000555398 | FLRT2 | chr14 | 86025652 | 6.78E-13 | 14:86039100-86039400 | -0.068501907 |
| ENST00000557419 | FLRT2 | chr14 | 85996506 | 3.54E-25 | 14:86039100-86039400 | -0.068501907 |
| ENST00000556335 | FLRT2 | chr14 | 85996489 | 4.37E-23 | 14:86039100-86039400 | -0.068501907 |
| ENST00000553627 | FLRT2 | chr14 | 86027206 | 0.000139792 | 14:86039100-86039400 | -0.068501907 |

| | | | | | | |
|---|---|---|---|---|---|---|
| ENST00000554746 | FLRT2 | chr14 | 85996571 | 3.92E-16 | 14:86039100-86039400 | -0.068501907 |
| ENST00000330753 | FLRT2 | chr14 | 85996487 | 3.68E-26 | 14:86039100-86039400 | -0.068501907 |
| ENST00000481272 | FSD1L | chr9 | 108210382 | 0.0690555 | 9:108210300-108210600 | -0.103080887 |
| ENST00000394926 | FSD1L | chr9 | 108210403 | 0.072740618 | 9:108210300-108210600 | -0.103080887 |
| ENST00000484973 | FSD1L | chr9 | 108210403 | 0.069549009 | 9:108210300-108210600 | -0.103080887 |
| ENST00000539376 | FSD1L | chr9 | 108210403 | 0.072740618 | 9:108210300-108210600 | -0.103080887 |
| ENST00000366714 | GJC2 | chr1 | 228337552 | 0.082825627 | 1:228336900-228337200 | -0.006508797 |
| ENST00000366714 | GJC2 | chr1 | 228337552 | 0.082825627 | 1:228337200-228337500 | -0.070049276 |
| ENST00000391865 | GUK1 | chr1 | 228327928 | 0.057289034 | 1:228336900-228337200 | -0.006508797 |
| ENST00000498092 | GUK1 | chr1 | 228328018 | 0.079428503 | 1:228336900-228337200 | -0.006508797 |
| ENST00000435153 | GUK1 | chr1 | 228328018 | 0.06358492 | 1:228336900-228337200 | -0.006508797 |
| ENST00000493138 | GUK1 | chr1 | 228328007 | 0.057879606 | 1:228336900-228337200 | -0.006508797 |
| ENST00000366721 | GUK1 | chr1 | 228328032 | 0.069753837 | 1:228336900-228337200 | -0.006508797 |
| ENST00000366718 | GUK1 | chr1 | 228332028 | 0.06457075 | 1:228336900-228337200 | -0.006508797 |
| ENST00000492871 | GUK1 | chr1 | 228327990 | 0.057242321 | 1:228337200-228337500 | -0.070049276 |
| ENST00000366730 | GUK1 | chr1 | 228327662 | 0.058130225 | 1:228336900-228337200 | -0.006508797 |
| ENST00000366726 | GUK1 | chr1 | 228327989 | 0.057289034 | 1:228336900-228337200 | -0.006508797 |
| ENST00000470040 | GUK1 | chr1 | 228332597 | 0.091467336 | 1:228336900-228337200 | -0.006508797 |
| ENST00000470155 | GUK1 | chr1 | 228333729 | 0.069741828 | 1:228336900-228337200 | -0.006508797 |
| ENST00000492871 | GUK1 | chr1 | 228327990 | 0.057242321 | 1:228336900-228337200 | -0.006508797 |

| | | | | | | |
|---|---|---|---|---|---|---|
| ENST00000366723 | GUK1 | chr1 | 228328002 | 0.058634059 | 1:228336900-228337200 | 0.006508797 |
| ENST00000485838 | GUK1 | chr1 | 228327966 | 0.057382535 | 1:228336900-228337200 | 0.006508797 |
| ENST00000312726 | GUK1 | chr1 | 228327990 | 0.057242321 | 1:228336900-228337200 | 0.006508797 |
| ENST00000453943 | GUK1 | chr1 | 228327998 | 0.099345026 | 1:228336900-228337200 | 0.006508797 |
| ENST00000366722 | GUK1 | chr1 | 228328017 | 0.06358492 | 1:228336900-228337200 | 0.006508797 |
| ENST00000366728 | GUK1 | chr1 | 228327990 | 0.057242321 | 1:228336900-228337200 | 0.006508797 |
| ENST00000366728 | GUK1 | chr1 | 228327990 | 0.057242321 | 1:228337200-228337500 | 0.070049276 |
| ENST00000366722 | GUK1 | chr1 | 228328017 | 0.06358492 | 1:228337200-228337500 | 0.070049276 |
| ENST00000366718 | GUK1 | chr1 | 228332028 | 0.06457075 | 1:228337200-228337500 | 0.070049276 |
| ENST00000366730 | GUK1 | chr1 | 228327662 | 0.058130225 | 1:228337200-228337500 | 0.070049276 |
| ENST00000498092 | GUK1 | chr1 | 228328018 | 0.079428503 | 1:228337200-228337500 | 0.070049276 |
| ENST00000312726 | GUK1 | chr1 | 228327990 | 0.057242321 | 1:228337200-228337500 | 0.070049276 |
| ENST00000453943 | GUK1 | chr1 | 228327998 | 0.099345026 | 1:228337200-228337500 | 0.070049276 |
| ENST00000366726 | GUK1 | chr1 | 228327989 | 0.057289034 | 1:228337200-228337500 | 0.070049276 |
| ENST00000366721 | GUK1 | chr1 | 228328032 | 0.069753837 | 1:228337200-228337500 | 0.070049276 |
| ENST00000470040 | GUK1 | chr1 | 228332597 | 0.091467336 | 1:228337200-228337500 | 0.070049276 |
| ENST00000435153 | GUK1 | chr1 | 228328018 | 0.06358492 | 1:228337200-228337500 | 0.070049276 |
| ENST00000391865 | GUK1 | chr1 | 228327928 | 0.057289034 | 1:228337200-228337500 | 0.070049276 |
| ENST00000366723 | GUK1 | chr1 | 228328002 | 0.058634059 | 1:228337200-228337500 | 0.070049276 |
| ENST00000493138 | GUK1 | chr1 | 228328007 | 0.057879606 | 1:228337200-228337500 | 0.070049276 |

| | | | | | | |
|---|---|---|---|---|---|---|
| ENST00000470155 | GUK1 | chr1 | 228333729 | 0.069741828 | 1:228337200-228337500 | 0.070049276 |
| ENST00000485838 | GUK1 | chr1 | 228327966 | 0.057382535 | 1:228337200-228337500 | 0.070049276 |
| ENST00000312841 | HR | chr8 | 21988565 | 0.001541921 | 8:21988800-21989100 | 0.070049276 |
| ENST00000522016 | HR | chr8 | 21975958 | 0.085114094 | 8:21988800-21989100 | 0.070049276 |
| ENST00000518461 | HR | chr8 | 21980203 | 0.015498577 | 8:21988800-21989100 | 0.070049276 |
| ENST00000381418 | HR | chr8 | 21989381 | 0.001687258 | 8:21988800-21989100 | 0.070049276 |
| ENST00000517699 | HR | chr8 | 21980364 | 0.029424519 | 8:21988800-21989100 | 0.070049276 |
| ENST00000422847 | LINC00857 | chr10 | 81967465 | 0.057879606 | 10:81956100-81956400 | 0.095911217 |
| ENST00000527924 | LTB4R2 | chr14 | 24774939 | 0.00099995 | 14:24728400-24728700 | 0.189913141 |
| ENST00000528054 | LTB4R2 | chr14 | 24778160 | 0.090965671 | 14:24728400-24728700 | 0.189913141 |
| ENST00000574453 | MINK1 | chr17 | 4736707 | 0.008155703 | 17:4699200-4699500 | 0.07173949 |
| ENST00000577021 | MINK1 | chr17 | 4736915 | 0.070013894 | 17:4699200-4699500 | 0.07173949 |
| ENST00000355280 | MINK1 | chr17 | 4736682 | 0.008209933 | 17:4699200-4699500 | 0.07173949 |
| ENST00000347992 | MINK1 | chr17 | 4736694 | 0.008190657 | 17:4699200-4699500 | 0.07173949 |
| ENST00000572330 | MINK1 | chr17 | 4736693 | 0.009253056 | 17:4699200-4699500 | 0.07173949 |
| ENST00000453408 | MINK1 | chr17 | 4736878 | 0.005383726 | 17:4699200-4699500 | 0.07173949 |
| ENST00000572330 | MINK1 | chr17 | 4736693 | 0.009253056 | 17:4752300-4752600 | 0.0834447 |
| ENST00000577021 | MINK1 | chr17 | 4736915 | 0.070013894 | 17:4752300-4752600 | 0.0834447 |
| ENST00000574453 | MINK1 | chr17 | 4736707 | 0.008155703 | 17:4752300-4752600 | 0.0834447 |
| ENST00000355280 | MINK1 | chr17 | 4736682 | 0.008209933 | 17:4752300-4752600 | 0.0834447 |

| | | | | | | |
|---|---|---|---|---|---|---|
| ENST00000453408 | MINK1 | chr17 | 4736878 | 0.005383726 | 17:4752300-4752600 | 0.0834447 |
| ENST00000347992 | MINK1 | chr17 | 4736694 | 0.008190657 | 17:4752300-4752600 | 0.0834447 |
| ENST00000572629 | MINK1 | chr17 | 4795765 | 0.025775589 | 17:4752300-4752600 | 0.0834447 |
| ENST00000533746 | NAV2 | chr11 | 20125756 | 0.014475262 | 11:20134200-20134500 | 0.096133386 |
| ENST00000267425 | NOP9 | chr14 | 24769067 | 0.000157684 | 14:24728400-24728700 | 0.189913141 |
| ENST00000557362 | NOP9 | chr14 | 24771610 | 0.000683345 | 14:24728400-24728700 | 0.189913141 |
| ENST00000510998 | PALLD | chr4 | 169633309 | 0.037687548 | 4:169615800-169616100 | 0.133019828 |
| ENST00000396039 | PARD6B | chr20 | 49348201 | 0.001494298 | 20:49345500-49345800 | 0.139441111 |
| ENST00000371610 | PARD6B | chr20 | 49348080 | 0.002214978 | 20:49345500-49345800 | 0.139441111 |
| ENST00000369349 | PDE4DIP | chr1 | 144994876 | 0.000274385 | 1:144966900-144967200 | 0.196195734 |
| ENST00000369354 | PDE4DIP | chr1 | 144994921 | 8.27E-05 | 1:144966900-144967200 | 0.196195734 |
| ENST00000313431 | PDE4DIP | chr1 | 144932365 | 1.38E-05 | 1:144966900-144967200 | 0.196195734 |
| ENST00000525886 | PDE4DIP | chr1 | 144919011 | 0.000180818 | 1:144966900-144967200 | 0.196195734 |
| ENST00000467859 | PDE4DIP | chr1 | 144921903 | 0.000283597 | 1:144966900-144967200 | 0.196195734 |
| ENST00000479408 | PDE4DIP | chr1 | 144932552 | 3.00E-05 | 1:144966900-144967200 | 0.196195734 |
| ENST00000529945 | PDE4DIP | chr1 | 144932148 | 4.59E-05 | 1:144966900-144967200 | 0.196195734 |
| ENST00000369356 | PDE4DIP | chr1 | 144995022 | 8.06E-05 | 1:144966900-144967200 | 0.196195734 |
| ENST00000369351 | PDE4DIP | chr1 | 144994840 | 0.000278318 | 1:144966900-144967200 | 0.196195734 |
| ENST00000377238 | PLXDC2 | chr10 | 20332248 | 0.045464866 | 10:20323800-20324100 | 0.028673864 |
| ENST00000414037 | PMEPA1 | chr20 | 56265680 | 8.09E-08 | 20:56232900-56233200 | 0.150476357 |

| | | | | | | |
|---|---|---|---|---|---|---|
| ENST00000395814 | PMEPA1 | chr20 | 56266886 | 2.67E-08 | 20:56232900-56233200 | -0.150476357 |
| ENST00000564977 | RP11-395B7.7 | chr7 | 100737327 | 0.001294333 | 7:100781400-100781700 | -0.005348181 |
| ENST00000223095 | SERPINE1 | chr7 | 100770369 | 1.82E-22 | 7:100781400-100781700 | -0.005348181 |
| ENST00000445463 | SERPINE1 | chr7 | 100770378 | 1.82E-22 | 7:100781400-100781700 | -0.005348181 |
| ENST00000227495 | ST3GAL4 | chr11 | 126225554 | 0.012763485 | 11:126264600-126264900 | -0.004513582 |
| ENST00000532243 | ST3GAL4 | chr11 | 126276089 | 0.081705961 | 11:126264600-126264900 | -0.004513582 |
| ENST00000530591 | ST3GAL4 | chr11 | 126225788 | 0.034006979 | 11:126264600-126264900 | -0.004513582 |
| ENST00000524834 | ST3GAL4 | chr11 | 126276397 | 0.060294765 | 11:126264600-126264900 | -0.004513582 |
| ENST00000531217 | ST3GAL4 | chr11 | 126225554 | 0.012763485 | 11:126264600-126264900 | -0.004513582 |
| ENST00000534083 | ST3GAL4 | chr11 | 126225868 | 0.015246956 | 11:126264600-126264900 | -0.004513582 |
| ENST00000356132 | ST3GAL4 | chr11 | 126225586 | 0.014190003 | 11:126264600-126264900 | -0.004513582 |
| ENST00000444328 | ST3GAL4 | chr11 | 126225578 | 0.014190003 | 11:126264600-126264900 | -0.004513582 |
| ENST00000449406 | ST3GAL4 | chr11 | 126276081 | 0.081705961 | 11:126264600-126264900 | -0.004513582 |
| ENST00000515433 | TGFBI | chr5 | 135381433 | 1.42E-11 | 5:135389100-135389400 | -0.179268968 |
| ENST00000514242 | TGFBI | chr5 | 135391187 | 7.72E-07 | 5:135389100-135389400 | -0.179268968 |
| ENST00000509485 | TGFBI | chr5 | 135389631 | 2.12E-06 | 5:135389100-135389400 | -0.179268968 |
| ENST00000514554 | TGFBI | chr5 | 135385204 | 1.32E-05 | 5:135389100-135389400 | -0.179268968 |
| ENST00000509749 | TGFBI | chr5 | 135383041 | 1.70E-09 | 5:135389100-135389400 | -0.179268968 |
| ENST00000504411 | TGFBI | chr5 | 135396878 | 0.050856475 | 5:135389100-135389400 | -0.179268968 |
| ENST00000504185 | TGFBI | chr5 | 135364676 | 1.31E-16 | 5:135389100-135389400 | -0.179268968 |

| | | | | | | |
|---|---|---|---|---|---|---|
| ENST00000604555 | TGFBI | chr5 | 135383106 | 2.97E-11 | 5:135389100-135389400 | 0.179268968 |
| ENST00000508076 | TGFBI | chr5 | 135394537 | 0.014737824 | 5:135389100-135389400 | 0.179268968 |
| ENST00000508767 | TGFBI | chr5 | 135383012 | 1.79E-10 | 5:135389100-135389400 | 0.179268968 |
| ENST00000503087 | TGFBI | chr5 | 135394874 | 0.001440823 | 5:135389100-135389400 | 0.179268968 |
| ENST00000305126 | TGFBI | chr5 | 135364690 | 1.48E-09 | 5:135389100-135389400 | 0.179268968 |
| ENST00000513497 | TGFBI | chr5 | 135394455 | 0.000154183 | 5:135389100-135389400 | 0.179268968 |
| ENST00000442011 | TGFBI | chr5 | 135364583 | 1.49E-09 | 5:135389100-135389400 | 0.179268968 |
| ENST00000506699 | TGFBI | chr5 | 135364679 | 1.48E-09 | 5:135389100-135389400 | 0.179268968 |
| ENST00000507018 | TGFBI | chr5 | 135364827 | 9.10E-10 | 5:135389100-135389400 | 0.179268968 |
| ENST00000206765 | TGM1 | chr14 | 24732416 | 0.021483222 | 14:24728400-24728700 | 0.189913141 |
| ENST00000559136 | TGM1 | chr14 | 24729364 | 0.007226709 | 14:24728400-24728700 | 0.189913141 |
| ENST00000544573 | TGM1 | chr14 | 24732386 | 0.021961714 | 14:24728400-24728700 | 0.189913141 |
| ENST00000399423 | TINF2 | chr14 | 24711869 | 0.003958201 | 14:24728400-24728700 | 0.189913141 |
| ENST00000540705 | TINF2 | chr14 | 24711811 | 0.00239228 | 14:24728400-24728700 | 0.189913141 |
| ENST00000267415 | TINF2 | chr14 | 24711880 | 0.003958201 | 14:24728400-24728700 | 0.189913141 |
| ENST00000558566 | TINF2 | chr14 | 24711819 | 0.003892034 | 14:24728400-24728700 | 0.189913141 |
| ENST00000558703 | TINF2 | chr14 | 24709448 | 0.028232225 | 14:24728400-24728700 | 0.189913141 |
| ENST00000560019 | TINF2 | chr14 | 24709680 | 0.029209997 | 14:24728400-24728700 | 0.189913141 |
| ENST00000559019 | TINF2 | chr14 | 24711838 | 0.040787474 | 14:24728400-24728700 | 0.189913141 |
| ENST00000558476 | TINF2 | chr14 | 24711791 | 0.062978587 | 14:24728400-24728700 | 0.189913141 |

| | | | | | | |
|---|---|---|---|---|---|---|
| ENST00000559969 | TINF2 | chr14 | 24711494 | 0.003022643 | 14:24728400-24728700 | -0.189913141 |
| ENST00000538777 | TINF2 | chr14 | 24711813 | 0.003234488 | 14:24728400-24728700 | -0.189913141 |
| ENST00000557915 | TINF2 | chr14 | 24709804 | 0.011601636 | 14:24728400-24728700 | -0.189913141 |
| ENST00000588410 | TPM4 | chr19 | 16187830 | 0.058130225 | 19:16181400-16181700 | -0.12166914 |
| ENST00000588507 | TPM4 | chr19 | 16186909 | 0.018035979 | 19:16181400-16181700 | -0.12166914 |
| ENST00000538887 | TPM4 | chr19 | 16178347 | 0.008325315 | 19:16181400-16181700 | -0.12166914 |
| ENST00000300933 | TPM4 | chr19 | 16187115 | 0.007672728 | 19:16181400-16181700 | -0.12166914 |
| ENST00000587201 | TPM4 | chr19 | 16187308 | 0.006910523 | 19:16181400-16181700 | -0.12166914 |
| ENST00000586833 | TPM4 | chr19 | 16186495 | 0.019342122 | 19:16181400-16181700 | -0.12166914 |
| ENST00000592138 | TPM4 | chr19 | 16192532 | 0.064559904 | 19:16181400-16181700 | -0.12166914 |
| ENST00000588032 | TPM4 | chr19 | 16198868 | 0.011490924 | 19:16181400-16181700 | -0.12166914 |
| ENST00000592822 | TPM4 | chr19 | 16204381 | 0.060650385 | 19:16181400-16181700 | -0.12166914 |
| ENST00000591226 | TPM4 | chr19 | 16210476 | 0.018963255 | 19:16181400-16181700 | -0.12166914 |
| ENST00000588483 | TPM4 | chr19 | 16191635 | 0.026846923 | 19:16181400-16181700 | -0.12166914 |
| ENST00000344824 | TPM4 | chr19 | 16178316 | 0.006673061 | 19:16181400-16181700 | -0.12166914 |
| ENST00000586499 | TPM4 | chr19 | 16178509 | 0.00807779 | 19:16181400-16181700 | -0.12166914 |
| ENST00000591645 | TPM4 | chr19 | 16203916 | 0.019926259 | 19:16181400-16181700 | -0.12166914 |
| ENST00000590180 | TPM4 | chr19 | 16190865 | 0.052168771 | 19:16181400-16181700 | -0.12166914 |
| ENST00000487252 | TRIM56 | chr7 | 100732891 | 0.040597794 | 7:100781400-100781700 | -0.005348181 |

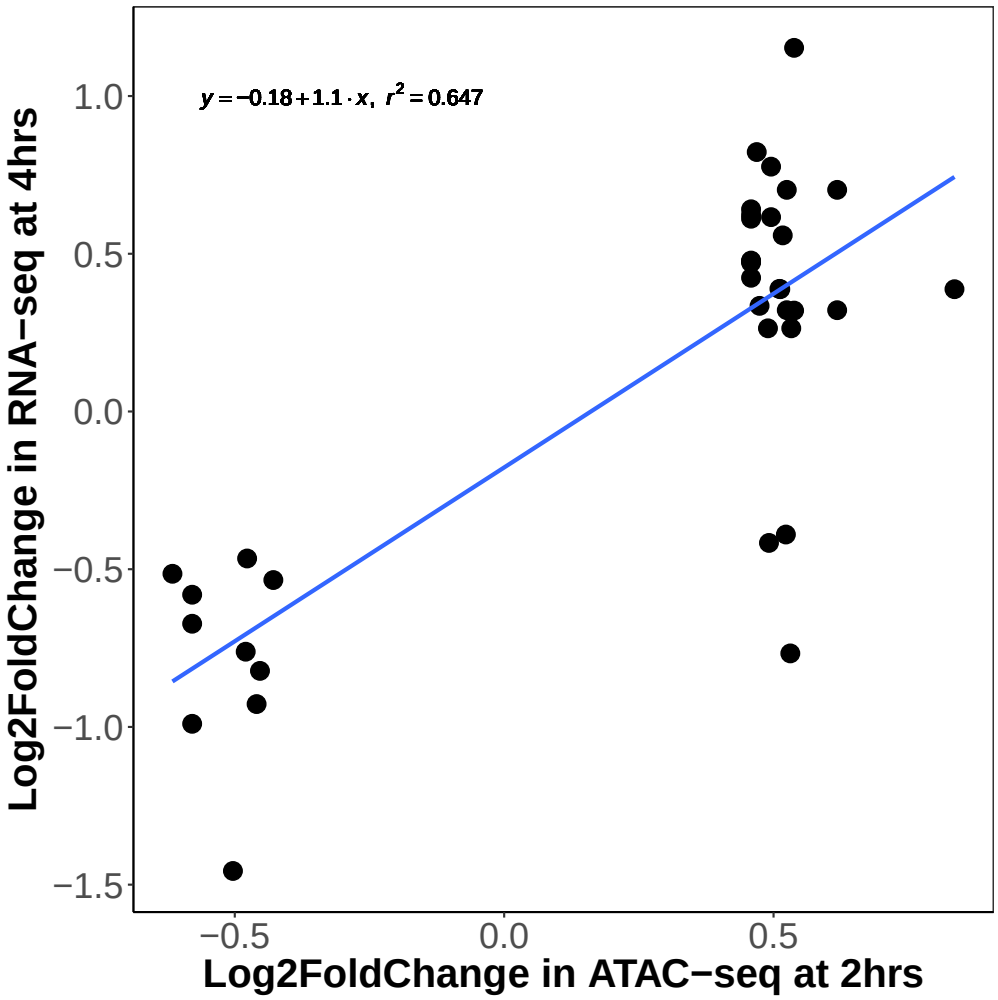


Figure S11: **Correlation of Changes in Chromatin Accessibility and Gene Expression.** Each point denotes an ATAC region and a gene within 50kb. Plot shows the correlation of changes in chromatin accessibility via ATAC-seq and changes in gene expression of a gene via RNA-seq ($r^2$= 0.647,p$-value$= 3.361 $\times 10^{-9}$).

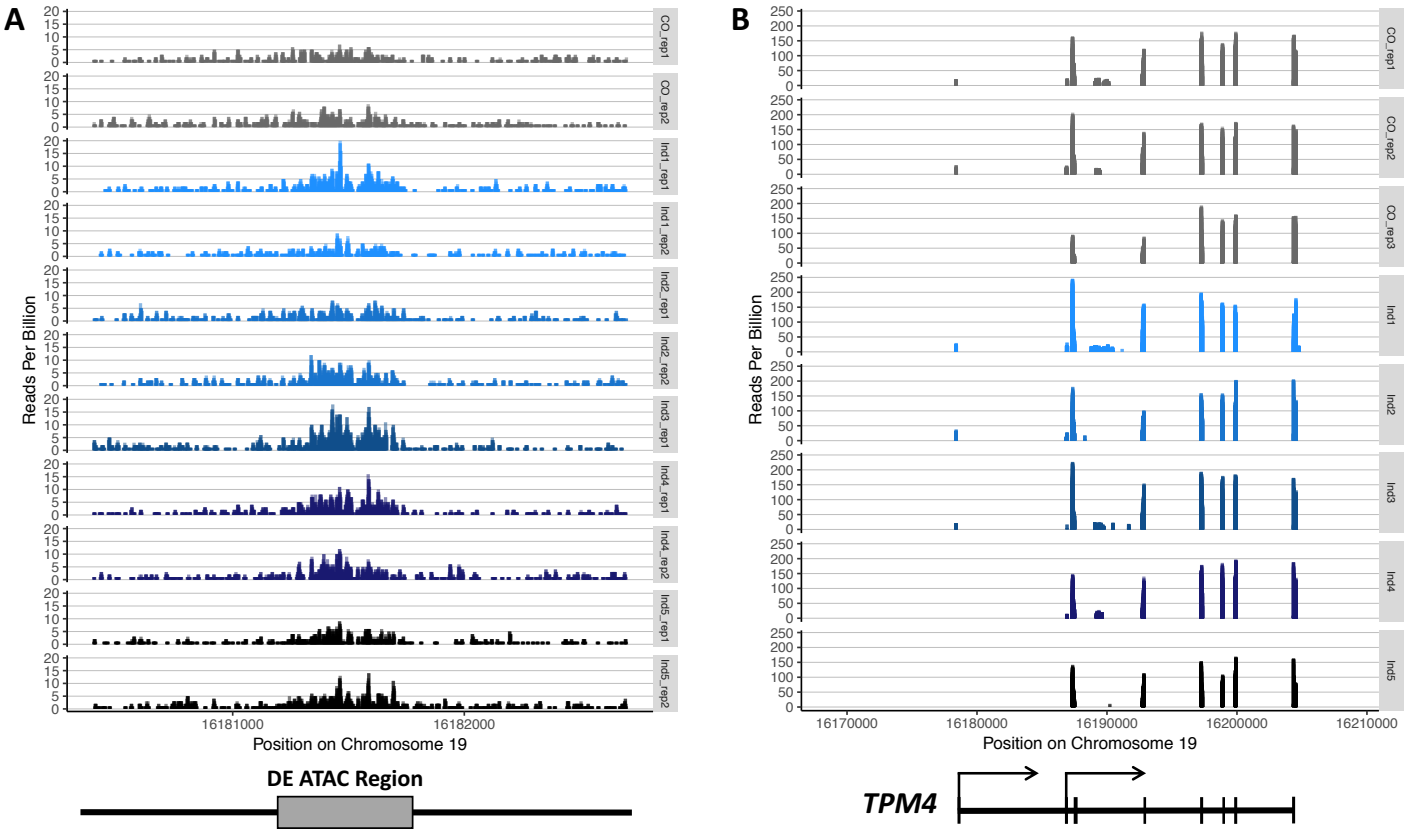


Figure S12: **ATAC and RNA-seq for TPM4 Separated by Sample. A** ATAC-seq profile centered on the 300bp window (with 1000bp on either side) that is differentially accessible following 2 hours of treatment with the microbiome (BH FDR = 12%). This region is found 4,795bp from the differentially expressed gene *TPM4*. **B** RNA-seq profile of *TPM4* (ENST00000586833) (and 20,000bp in either direction) which is differentially expressed following 4 hours of treatment with the microbiome (BH FDR = 1.9%). The gene model is shown below. The tracks in **A** and **B** are equivalent to the averaged tracks in Figure 4B and C, except that here they are separated by sample with each color denoting a different treatment. Replicates are separate but are coded in the same color (gray as control).

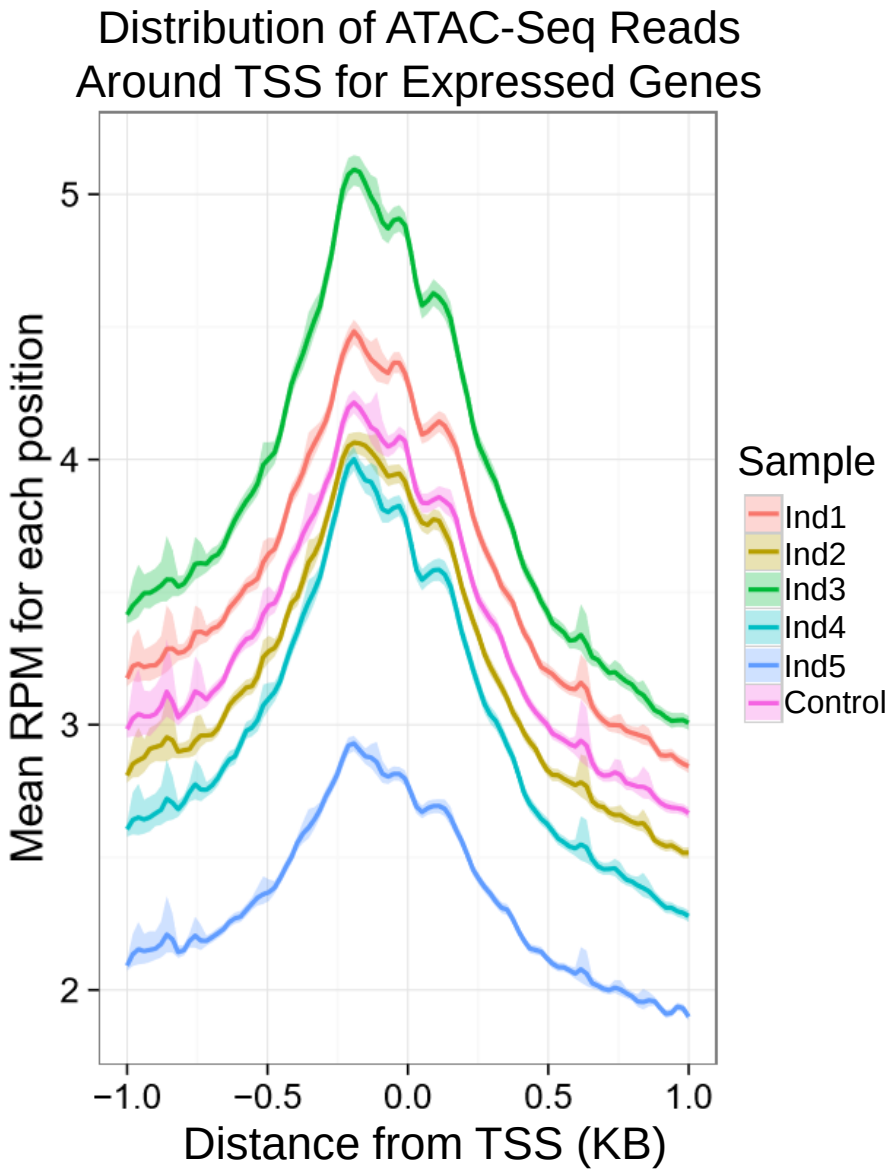


Figure S13: **Distribution of ATAC-seq reads around gene transcriptional start sites (TSS).** For each of the five samples and the control, we show the average number of ATAC-seq reads per million (RPM) at each base position within 1kb upstream and downstream of the transcriptional start site for all genes expressed at 4 hours. This plot was generated with metagene [20].

Table S12: **Active Motifs in Treatment vs Control Conditions.** We identified 40 motifs with footprints in regions differentially accessible including factors that bind the ELF5, SOX2, and HMGA2 motifs. Motif enrichment or depletion in differentially accessible regions (DARs) was calculated using the Fisher's Exact test for each motif. Each row of the table represents a significantly enriched or depleted motif. Columns include: Motif = name of motif; Odds_Ratio = odds ratio of enrichment or depletion; Lower = lower bound of 95% confidence interval; Upper = upper bound of 95% confidence interval; Enrichment.Pvalue = $p$-value of odds ratio; TFpos_DARpos = number of DARs containing footprint; TFpos_DARneg = number of non-differentially accessible regions containing footprint; TFneg_DARpos = number of DARs not containing footprint; TFneg_DARneg = number of non-differentially accessible regions not containing footprint; Treatment_foots = total number of footprints in treatment condition; Control_foots = total number of footprints in control condition

(see TableS12.txt)

Table S13: **Differentially Accessible Regions by Stratified FDR.** Table with 209 differentially accessible regions (DARs) from stratified FDR test on ATAC-seq (BH FDR $< 0.1$). Each row of the table represents a DAR. Columns include: Chr = chromosome; Start = beginning of DAR; End = end of DAR; baseMean = mean of normalized ATAC-seq read counts across all samples, normalized for sequencing depth; log2FC = $\log_2$-transformed fold change; lfcSE = standard error of log fold change; stat = Wald statistic (log2FC / lfcSE); p.val = $p$-value of region being differentially accessible; DESeq_padj = non-stratified FDR adjustment; motifs = list of all motifs with footprint in region.
(see TableS13.txt)

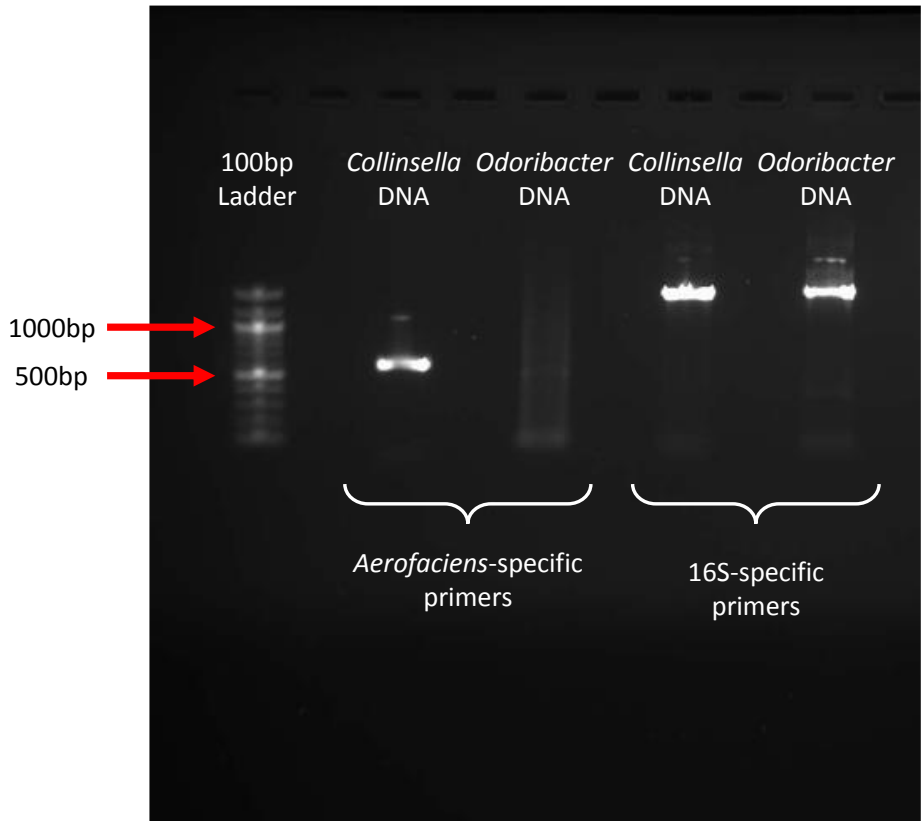


Figure S14: **Confirmation of *Collinsella aerofaciens* DNA.** Gel with five wells: 100bp ladder, *Collinsella aerofaciens* DNA and *Odoribacter splanchnicus* DNA with primers specific to 16S region of *Collinsella aerofaciens* (590bp), and *Collinsella aerofaciens* DNA and *Odoribacter splanchnicus* DNA with primers to amplify 16S region in any microbial sample.

# References


[1] Richards, A.L., Burns, M.B., Alazizi, A., Barreiro, L.B., Pique-Regi, R., Blekhman, R., Luca, F.: Genetic and transcriptional analysis of human host response to healthy gut microbiota. mSystems **1**(4) (2016)

[2] Kageyama, A., Sakamoto, M., Benno, Y.: Rapid identification and quantification of Collinsella aerofaciens using PCR. FEMS Microbiol. Lett. **183**(1), 43–47 (2000)

[3] Moyerbrailean, G.A., Davis, G.O., Harvey, C.T., Watza, D., Wen, X., Pique-Regi, R., Luca, F.: A high-throughput RNA-seq approach to profile transcriptional responses. Sci Rep **5**, 14976 (2015)

[4] Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., Gingeras, T.R.: STAR: ultrafast universal RNA-seq aligner. Bioinformatics **29**(1), 15–21 (2013)

[5] Love, M.I., Huber, W., Anders, S.: Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. **15**(12), 550 (2014)

[6] Backes, C., Keller, A., Kuentzer, J., Kneissl, B., Comtesse, N., Elnakady, Y.A., Muller, R., Meese, E., Lenhof, H.P.: GeneTrail–advanced gene set enrichment analysis. Nucleic Acids Res. **35**(Web Server issue), 186–192 (2007)

[7] Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., Klemm, A., Flicek, P., Manolio, T., Hindorff, L., Parkinson, H.: The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res. **42**(Database issue), 1001–1006 (2014)

[8] Burns, M.B., Lynch, J., Starr, T.K., Knights, D., Blekhman, R.: Virulence genes are a signature of the microbiome in the colorectal tumor microenvironment. Genome Med **7**(1), 55 (2015)

[9] 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., Knight, R.: QIIME allows analysis of high-throughput community sequencing data. Nat. Methods **7**(5), 335–336 (2010)

[10] 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., Knight, R.: Advancing our understanding of the human microbiome using QIIME. Meth. Enzymol. **531**, 371–444 (2013)

[11] DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., Huber, T., Dalevi, D., Hu, P., Andersen, G.L.: Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl. Environ. Microbiol. **72**(7), 5069–5072 (2006)

[12] 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., White, O.: A framework for human microbiome research. Nature **486**(7402), 215–221 (2012)

[13] 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., White, O.: Structure, function and diversity of the healthy human microbiome. Nature **486**(7402), 207–214 (2012)

[14] Chassaing, B., Darfeuille-Michaud, A.: The commensal microbiota and enteropathogens in the pathogenesis of inflammatory bowel diseases. Gastroenterology **140**(6), 1720–1728 (2011)

[15] Hansen, R., Russell, R.K., Reiff, C., Louis, P., McIntosh, F., Berry, S.H., Mukhopadhya, 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., Hold, G.L.: Microbiota of de-novo pediatric IBD: increased Faecalibacterium prausnitzii and reduced bacterial diversity in Crohn's but not in ulcerative colitis. Am. J. Gastroenterol. **107**(12), 1913–1922 (2012)

[16] de Oliveira, G.L.V., Leite, A.Z., Higuchi, B.S., Gonzaga, M.I., Mariano, V.S.: Intestinal dysbiosis and probiotic applications in autoimmune diseases. Immunology (2017)

[17] Lin, L., Zhang, J.: Role of intestinal microbiota and metabolites on gut homeostasis and human diseases. BMC Immunol. **18**(1), 2 (2017)

[18] Turnbaugh, P.J., Backhed, F., Fulton, L., Gordon, J.I.: Diet-induced obesity is linked to marked but reversible alterations in the mouse distal gut microbiome. Cell Host Microbe **3**(4), 213–223 (2008)

[19] Friedman, J., Alm, E.J.: Inferring correlation networks from genomic survey data. PLoS Comput. Biol. **8**(9), 1002687 (2012)

[20] Joly Beauparlant, C., Lamaze, F.C., Deschenes, A., Samb, R., Lemacon, A., Belleau, P., Bilodeau, S., Droit, A.: metagene Profiles Analyses Reveal Regulatory Element's Factor-Specific Recruitment Patterns. PLoS Comput. Biol. **12**(8), 1004751 (2016)