Development of Inflammatory Bowel Disease is Linked to a Longitudinal Restructuring of the Gut Metagenome in Mice.

Thomas Sharpton[1,2]*, Svetlana Lyalina[3]*, Julie Luong[3], Joey Pham[3], Emily M. Deal[3], Courtney Armour[1], Christopher Gaulke[1], Shomyseh Sanjabi[3,4], Katherine S. Pollard[3,5#]

[1] Department of Microbiology, Oregon State University, Corvallis, OR, USA

[2] Department of Statistics, Oregon State University, Corvallis, OR, USA

[3] Gladstone Institutes, San Francisco, CA, USA

[4] Department of Microbiology & Immunology, University of California, San Francisco, CA, USA

[5] Department of Epidemiology & Biostatistics, Institute for Human Genetics, and Institute for Computational Health Sciences, University of California, San Francisco, CA, USA

* These authors contributed equally to this work

[#] Corresponding author

E-mail: kpollard@gladstone.ucsf.edu

Running title: Longitudinal Mouse IBD Microbiome

Abstract

The gut microbiome is linked to inflammatory bowel disease (IBD) severity and altered in late stage disease. However, it is unclear how gut microbial communities change over the course of IBD development, especially in regards to function. To investigate microbiome mediated disease mechanisms and discover early biomarkers of IBD, we conducted a longitudinal metagenomic investigation in an established mouse model of IBD, where dampened TGF-β signaling in T cells leads to peripheral immune activation, weight loss, and severe colitis. IBD development is associated with abnormal gut microbiome temporal dynamics, including dampened acquisition of functional diversity and significant differences in abundance trajectories for KEGG modules such as glycosaminoglycan degradation, cellular chemotaxis, and type III and IV secretion systems. Most differences between sick and control mice emerge when mice begin to lose weight and heightened T cell activation is detected in peripheral blood. However, lipooligosaccharide transporter abundance diverges prior to immune activation, indicating that it could be a pre-disease indicator or microbiome-mediated disease mechanism. Taxonomic structure of the gut microbiome also significantly changes in association with IBD development, and the abundance of particular taxa, including several species of *Bacteroides*, correlate with immune activation. These discoveries were enabled by our use of generalized linear mixed effects models to test for differences in longitudinal profiles between healthy and diseased mice while accounting for the distributions of taxon and gene counts in metagenomic data. These findings demonstrate that longitudinal metagenomics is useful for discovering potential mechanisms through which the gut microbiome becomes altered in IBD.

Importance

IBD patients harbor distinct microbial communities with different functional capabilities compared to healthy people. But is this cause or effect? Answering this question requires data on changes in gut microbial communities leading up to disease onset. By performing weekly metagenomic sequencing and mixed effects modeling on an established mouse model of IBD, we identified several functional pathways encoded by the gut microbiome that covary with host immune status. These pathways are novel early biomarkers that may either enable microbes to live inside an inflamed gut or contribute to immune activation in IBD mice. Future work will validate the potential roles of these microbial pathways in host-microbe interactions and human disease. This study is novel in its longitudinal design and focus on microbial pathways, which provided new mechanistic insights into the role of gut microbes in IBD development.

Keywords: gut microbiome, metagenome, inflammatory bowel disease, mouse model, TGF-beta (β), autoimmunity, longitudinal modeling, biomarkers, T cell activation

Background

Inflammatory bowel disease (IBD) is an increasingly prevalent chronic autoimmune disease wherein the cells of the immune system attack intestinal tissue [1-3]. Quality of life deteriorates, and patients die in severe cases. Unfortunately, the etiology of disease remains unclear and is likely complex [4]. Discovery of the factors that contribute to IBD onset, development, and severity is needed to ensure accurate and effective health care. Epidemiological studies and animal model experiments have identified genetic [5-7] and lifestyle factors that associate with IBD, including diet [8] and exercise [9]. But these factors are not precise predictors of disease risk, severity, or response to treatment, and many questions remain regarding disease mechanisms. Elucidating the cryptic etiology of IBD would enable new preventative measures, diagnostics, and therapies.

Recent work has implicated the gut microbiome in the development and severity of IBD [10]. Individuals afflicted with Crohn's disease or ulcerative colitis, the two principal clinical forms of IBD, harbor distinct taxa relative to healthy controls [11-14]. Shotgun metagenomics further revealed that the abundance of several microbial metabolic pathways are significantly altered in IBD guts [13, 15, 16]. These associations may be causal, because gut microbes can influence the immune system and intestinal homeostasis. For example, immunosuppressive regulatory T cells (Tregs) are prevalent in the colonic lamina propria (LP) compared to other organs. But, their numbers are reduced in germ-free or antibiotic-treated mice, suggesting that microbiota affect colonic differentiation of peripheral Tregs (pTregs) [17, 18]. A similar loss of Tregs occurs in people with polymorphisms in IBD-susceptibility genes that promote defects in Treg responses [19]. Thus, gut microbes have the potential to interact with immune cells and this interaction can be altered due to host genetics and other risk factors in the development of IBD.

We hypothesized that the changes in immune status of individuals with IBD are associated with temporal alterations in the functional capabilities of their gut microbiota. Understanding how the gut

microbiome dynamically changes during IBD and how these changes relate to host symptoms and immune activation could clarify which microbiomic alterations contribute to disease onset and progression and which alterations respond to disease. We are particularly interested in elucidating specific microbial pathways that may induce or exacerbate immune activation and distinguishing these from pathways required for survival in an inflamed intestinal environment. Addressing these questions requires a prospective, longitudinal study of the microbiome in IBD.

Longitudinal investigations of the microbiome have tended to focus on taxonomic rather than functional changes [20, 21]. One study used 16S sequencing in the T-bet-/- RAG2-/- Ulcerative Colitis (TRUC) mouse model of inflammatory disease to identify how gut microbiome taxonomic composition changes over the course of treatment-induced remission and then imputed how microbial pathway abundances might change over time with ancestral state reconstruction techniques [22]. Shotgun metagenomic sequencing provides direct insight into the functions encoded in the microbiome, but it has not been applied to a longitudinal investigation of IBD. As a result, our insight into how the gut microbiome operates dynamically in association with disease development is limited.

Mouse models of disease present an opportunity to quantify the longitudinal covariation between gut microbiome functions and IBD development while overcoming the challenges associated with a prospective human study and reducing the extensive genetic, lifestyle, and microbiome variation among humans. We implemented this approach using a well-documented IBD model [23-29], where TGF-β dominant negative receptor II is driven by the CD4 promoter (CD4-dnTβRII) [30], called DNR hereafter. TGF-β is important for inducing pTreg differentiation [31], and its signaling in naive T cells results in activation and nuclear translocation of Smad2/3 molecules and regulation of target genes, including Foxp3 [32-34]. Foxp3 then provides a positive feedback loop by downregulating Smad7, thereby reducing its inhibition of TGF-β signaling [35]. Absence of TGF-β signaling in T cells results in loss of Foxp3 expression and defective *in vivo* expansion and immunosuppressive capacity of pTregs [36, 37]. However, excess inflammation can also potently inhibit Foxp3 induction by TGF-β [38, 39], and the

presence of certain inflammatory cytokines can instead divert differentiation of Tregs into pathogenic Th17 cells [40-45]. Thus, due to TGF-β's involvement in Treg cell differentiation, and the requirement for Treg produced IL-10 to maintain intestinal homeostasis, TGF-β signaling in T cells is an important component of intestinal immunity [46-54]. Furthermore, mutations in both TGF-β and IL-10 signaling pathways have been implicated in human IBD [55-58]. As a result of the blockage of TFGβ signaling on their T cells, and reduced number of pTregs, DNR animals develop spontaneous colonic inflammation and IBD that is akin to Crohn's disease [30, 59]. In addition to these physiological similarities, the DNR line serves as an effective model of human IBD because (i) human IBD is associated with mutations in SMAD3 [5, 60-62], a direct downstream target of TGFβ RII required for Foxp3 induction in the gut [33], and (ii) DNR mice model the documented effect of Smad7 overexpression in human IBD [63-65].

To obtain insight into how the longitudinal dynamics of the microbiome associate with IBD onset and progression, we followed DNR and littermate wild-type (hereafter, WT) controls from weaning through severe disease. We used shotgun metagenomics to quantify how fecal microbiome structure and function change over the course of disease development in DNR mice and identified components of the microbiome that both associate with and predict immune status. We focus on longitudinal changes in biological pathways (i.e., groups of genes performing a coherent function), using estimated abundances of KEGG modules from DNR and WT metagenomes. Our work indicates that the microbiome may contain biomarkers of IBD development, clarifies mechanisms through which the microbiome may contribute to disease development, and reveals how gut microbes operate to succeed in an inflamed intestinal environment.

Results

Age-matched female WT and DNR littermates were monitored longitudinally for IBD development over a period of 9 weeks, starting at 4 weeks of age upon being weaned from their mother. As this is a T cell-mediated IBD model, we quantified peripheral CD4 and CD8 T cell activation by flow cytometry and

measured the longitudinal change in the CD44^{hi} activated fraction, which includes both effector and memory T cells (Additional File 1: Figure S1). We also measured the weight of the animals over time (Fig. 1A). As expected, WT mice gained weight and maintained a constant fraction of activated T cells. DNR mice, conversely, stopped gaining weight and experienced a sharp increase in CD4 T cell activation followed by gradual increase in CD8 T cell activation starting at 7 weeks of age (Fig. 1). These results indicate that in our facility, the DNR mice develop signs of IBD starting around week 7 and full disease by week 9. DNR mice had to be euthanized by week 15, as they had lost more than 15% of their maximum body weight. Similar to the T cell activation phenotype observed in the blood after week 7 (Fig. 1B-C), the DNR animals had a larger fraction of activated T cells in the spleen and the gut-draining mesenteric lymph node (MLN) at week 15 (Additional File 2: Figure S2).

We used shotgun metagenomics to assess how the functional potential of the gut microbiome diversifies over the course of disease progression. Specifically, we collected stool samples from parallel cohorts of DNR and WT mice weekly and performed shotgun metagenomic sequencing from samples obtained at 4, 5, 6, 8, 10, 12, and 13 weeks of age (Additional File 3: Table S1). We then quantified the abundance of KEGG modules encoded in each metagenome with ShotMAP [16], which revealed 373 modules present in at least one sample. These module abundances were then used to quantify how the within-sample (alpha) diversity of microbiome functions varies over time in DNR and WT mice. A Kruskal-Wallis test of the change in KEGG module Shannon entropy over time (Additional File 4: Figure S3) found that the DNR mice are relatively stable in their functional alpha-diversity ($p=0.47$) as compared to WT mice ($p=0.078$). We also observed that functional alpha-diversity varies among individuals within a line over time, and that this variation differs between lines in association with disease activation (week 7). Specifically, the coefficient of variation of KEGG module Shannon entropy (CV) from a generalized linear model is higher among WT than DNR mice after disease activation ($p = 0.0085$). We also find that the CV is higher among DNR mice prior to activation, though this difference is reduced when the disproportionately variable week 5 samples are removed from the analysis

(p=0.21). These results show that the functional diversity of the mouse gut microbiome is relatively constrained early in life but increases over the lifetimes of WT but not DNR individuals.

We then investigated how the composition of gut microbiome functions varies over time and between cohorts (DNR vs. WT) by using an abundance-weighted beta-diversity metric (Bray-Curtis dissimilarity). At a global level, KEGG module abundances were similar between DNR and WT mice prior to week 6, but then diverged over time as IBD developed in the DNR mice (Fig. 2). Furthermore, the diversity of KEGG modules found in a metagenome was significantly associated with the week that the sample was collected within the cohort (PERMANOVA p=0.01, R^2=0.42), as well as cohort's weekly mean activated T cell status (pcCD4tCD44hi, PERMANOVA p=0.001, R^2=0.16). Thus, there exist microbiome-encoded functional modules that differ in abundance in association with IBD progression in DNR mice.

This temporal divergence in DNR versus WT microbiome functions is mirrored in the taxonomic structure of the microbiome (Fig. 2). The composition of the gut metagenomes is relatively similar between WT and DNR lines at early time points and begins to diverge at week 6. Additionally, the microbiomes of WT mice remain relatively consistent over time as compared to DNR mice, though they are not without temporal variation. Indeed, similar to the functional diversity analysis, the taxonomic beta-diversity of the microbiome significantly differs between the lines over time (PERMANOVA p=0.004, R^2=0.46), though not with mean activated T cell status (PERMANOVA p=0.118, R^2=0.046). Collectively, these analyses indicate that (1) the diversity and structure of the gut microbiome varies over time between 4 and 15 weeks of age in both WT and DNR mice, (2) WT and DNR microbiomes are generally consistent prior to immune activation in DNR mice, but diverge afterwards, and (3) immune activation is associated with changes in the subsequent succession of the gut microbiome.

Based on these observations, we assessed how specific components of the microbiome associate with disease development. A key novelty of our approach is the use of Tweedie compound Poisson generalized linear mixed effects models (GLMMs). These models allow us to test for differences in

188 temporal trends in KEGG module abundance between DNR and WT mice while accounting for baseline
189 differences between mice and genotypes, as well as DNA extraction kit effects. GLMMs enable
190 accurate modeling of non-normally distributed abundance data and correctly account for multiple
191 sources of variation [66], including the inter-subject variation that is present in repeated measures
192 designs such as the longitudinal sampling of individual mice in our study. The Tweedie compound
193 Poisson distribution, which is a weighted mixture between Poisson and Gamma distributions, has a
194 number of other attractive features. Its exponential relationship between variance and mean captures
195 the overdispersion that is frequently present in environmental DNA sequence data, and its point mass
196 at zero allows for one-step fitting of zero-inflated data (versus fitting a model to determine feature
197 presence/absence before modeling non-zero components, as in hurdle models). Additionally, the
198 Tweedie compound Poisson is a continuous distribution, allowing us to use a normalized abundance
199 measure as the dependent variable, instead of raw counts. We provide a more detailed description of
200 the models used in our analysis in Additional Data File 5: Text S1.
201
202 We first looked at overall trends of abundance trajectories for DNR versus WT mice as quantified by the
203 interaction between genotype and time in the GLMM. These analyses revealed 29 KEGG modules with
204 significant differences in abundance trends between DNR and WT mice (FDR<0.05). The interaction
205 coefficient was positive for 26 of the significant modules (Additional File 6: Table S2), which indicates
206 that these modules became increasingly abundant in DNR versus WT mice over time. This set includes
207 modules associated with uridine monophosphate biosynthesis (M00051), keratin sulfate degradation
208 (M00079), and the type III secretion system (M00332). The three modules with negative interaction
209 coefficients, indicating decreasing abundance in DNR versus WT mice over time (Fig. 3), are lysine
210 biosynthesis (M00031), lipooligosaccharide transport (M00252), and melatonin biosynthesis (M00037).
211
212 To obtain improved temporal resolution regarding the divergence of module abundance in DNR mice,
213 we extended our GLMMs to include a “hinge” at week 7, which is when immune activation initiates in
214 DNR mice. This segmented regression approach has the potential to discover modules that diverge in

abundance between DNR and WT mice either between weeks 4 and 7 or between weeks 7 and 13. Only 13 of the 29 previously identified modules exhibited a significant effect when using segmented regression (Fig. 4), likely due to a loss of power from partitioning the data into two smaller sets of samples. However for these 13 modules, our results clarify when DNR and WT abundances began to diverge (Additional File 7: Table S3). The predominant pattern was similar module abundance prior to week 7, followed by divergence after immune activation (11/13 modules). This pattern suggests that these modules respond to disease or play a role in disease progression.

Lipooligosaccharide transport (M00252), which is a two-component system with an unknown substrate in the mammalian gut, was the only module that stratified DNR and WT mice both before and after disease onset. To further investigate the potential taxa that may drive this particular signal, we assessed the taxonomic source of the KEGG sequences that recruited metagenomic reads into the module. We also quantified the distance covariance [67] between the longitudinal trajectories of the KEGG Orthology Groups (KOs) that comprise the module and each observed species' trajectory. The result was mixed, with the former analysis suggesting primarily *Streptococcus* contributions, while the latter identified greatest similarity with *Lactobacillus murinus* and *Candidatus* Arthromitus trajectories (Additional File 8: Figure S4). The differences in the taxonomic composition of the reference data underlying these two approaches could account for these inconsistencies, as could the fact that the KEGG analysis relies on amino acid comparisons while the species trajectories are determined through nucleotide comparisons. Thus, an uncharacterized lipooligosaccharide transporter encoded in *Streptococcus* and other gut microbes decreases in abundance over time at a significantly faster rate in DNR compared to WT mice, starting early in life before weight loss and immune activation.

Type III secretion system (M00332) differed in its temporal change between the lines uniquely before disease onset. Specifically, the module decreased in abundance in WT mice over weeks 1-7, with KO K03225 primarily driving this effect. On the other hand, this module was relatively stable in DNR mice prior to disease onset, and several of the KOs that comprise the module increased in DNR mice in the

later weeks (Fig. 4). The discovery of stable, rather than decreasing, abundance of K03225 as an early indicator of IBD in DNR mice is intriguing because Type III secretion systems are used by pathogens to invade the gut community and alter the gut environment [68, 69].

We next examined baseline differences in module abundance between DNR and WT mice at weaning. Early differences could result from genotype-specific selection of the gut microbiome or cage effects. Our models revealed 17 modules with significantly different intercepts ($q<0.05$), which indicates differences in abundance between the two lines at week 4 (Additional File 9: Table S4). Eight of these modules, including several methanogenesis associated pathways, had positive intercept coefficients, meaning that they were more abundant in DNR compared to WT mice at week 4. Lipopolysaccharide biosynthesis and eight other modules showed the opposite effect and were higher in WT mice at weaning. This early-life variation in the microbiome supports hypotheses that pre-adolescent development of the microbiome can affect health state later in life. However, these temporal relationships are complex: later changes in abundance, as captured by the time by cohort interaction, could reverse the pattern seen at weaning.

To explore temporal dynamics of specific gut taxa, we applied the same GLMM analysis to species abundances. This analysis yielded no significant results at FDR < 0.05, likely due to not having the advantage of grouping components across a higher order variable. While species could be grouped into higher taxonomic entities, the model assumption that members of the same group tend to covary across samples and over time may be violated because members of the same taxonomy may compete or ecologically exclude one another [70]. We evaluated this possibility by applying a non-parametric decomposition of variance components [71] to assess whether within-module or within-genus dispersion decomposition patterns were significantly different from those obtained from random permutations of the underlying data. This auxiliary analysis finds that components of functional groups covary more than random while components of taxonomic groups do not (Additional Data File 10: Figure S5). This observation indicates that grouping taxa would violate GLMM model assumptions.

Consequently, we instead used a goodness-of-fit test based on functional principal components analysis (FPCA), which is less rigid in its assumption of linearity and capable of borrowing information across species due to the representation of abundance trajectories as combinations of eigen-functions derived from the entire dataset. This test identified seven species that significantly differ in their variation over time between the DNR and WT cohorts (Figure 5; Table 1), including greater increases in abundance over time within DNR microbiomes for *E. coli* and four species from the *Bacteroides* genus, which are associated with gut inflammation [10].

Discussion

This study represents the first shotgun metagenomic characterization of IBD development. By using a controlled mouse model, a longitudinal study design, and statistical modeling, we identified novel microbial biomarkers associated with IBD onset and progression. Many of the taxa and functions we implicated have known roles in immune regulation and pathogenicity, making them plausible candidates for stimulating the disease process, while others likely represent responses of the microbiota to changes in host physiology. Ordination and GLMM analyses enabled us to distinguish these scenarios by identifying significant differences between DNR and WT mice over time from weaning through severe disease. We discovered that lipooligosaccharide transport and type III secretion protein abundance trajectories between weaning and immune activation differentiate DNR mice prior to immune activation, making them promising early biomarkers and consistent with a potentially causal role in IBD. Abundances of 17 modules are altered in DNR mice at weaning and could predict IBD risk if they generalize to other mouse models and human disease (see below). Many other modules as well as a few species have altered abundances in DNR mice in later, more severe stages of disease. Functional and taxonomic diversity also show temporal differences in DNR mice that correlate with immune profiles and/or disease progression. Most of these discoveries would have been missed in a cross-sectional study because the disease association is a longitudinal trend.

By using shotgun metagenomics, we were able to investigate both taxonomic and functional characteristics of the IBD microbiome. Both types of data consistently showed differences between DNR and WT mice. For example, beta-diversity analyses revealed increasing divergence of both taxonomic and functional profiles between DNR and WT microbiomes over the last four weeks of the study. In addition, the individual taxa and modules with genotype-specific trajectories predominantly had increased abundance in DNR mice after disease onset. These similarities in the successional diversification of species and genes support the idea that taxonomic changes in IBD have functional consequences that are linked to immune activation. Despite such parallels, our taxonomic and functional results differed in several important ways. Notably, a smaller number of species stratify lines over time as compared to KEGG modules. Furthermore, most of the IBD-associated modules we discovered were not represented solely in singular species and would have been missed by considering information from taxonomic analyses only. These results could be due to disease-associated functional redundancy, wherein a gene that is enriched in DNRs might derive from different species in each mouse. Other potential reasons include (1) higher power due to grouping protein families into modules, and (2) missed taxonomic associations due to the relatively small number of laboratory mouse associated microbes in the genome tree of life [72]. Future work should explore how taxa missed by reference-based quantification vary in association with IBD in DNR mice.

Despite finding relatively few species that distinguish DNR mice, we can gain insight into the disease process from what is known about how these taxa interact with the host. It is striking that four of the seven species that change in abundance as IBD develops belong to the genus *Bacteroides,* three of which are more abundant in DNR mice. Several studies have implicated *Bacteroides* in intestinal inflammation. For example, a subset of *B. fragilis* strains carry a proinflammatory metalloprotease toxin that has been identified in 19.3% of patients with active IBD [73], and the inoculation of animals with such strains is associated with severe colitis [10, 74]. Subsequent research showed that multiple commensal species of *Bacteroides* could be incorporated into the gut microbiomes of IBD-susceptible genotypes of mice to induce IBD, including mice with TGF-β susceptibility loci [75]. Supporting the idea

that *Bacteroides* contribute to IBD, we observe a modest increase (q=0.1898) in the hemophore/metalloprotease transport system module (M00328) in DNR mice as disease progresses. These and other mechanistic hypotheses must be tested, because the species of *Bacteroides* we identified are diverse and species within the same genus can exhibit discordant patterns of interaction with host physiology [76].

Cross-sectional and mechanistic investigations of IBD support our finding that disease development is linked to microbiome taxonomy and function [4, 77, 78]. The progressive divergence of DNR and WT microbiomes as IBD worsens is consistent with a 16S-based study using a different mouse model of IBD in which gut microbes and imputed functions changed in association with disease status and therapeutically induced remission [22]. Additionally, studies in germ-free mouse models of IBD implicate the gut microbiome in disease development. For example, interleukin (IL)-10-knockout mice grown under germ-free conditions do not develop colitis, while conventionally raised mice do [79]. Similar findings have been reported for the TRUC mouse model [80]. Furthermore, IL-10-knockout [81] and IL-2-deficient [82] mice manifest differential severity of colitis dependent on the types of taxa that colonize their gut. Human studies of IBD have yet to investigate the disease longitudinally. However, our results are consistent with microbiome case-control studies that found significant differences in the taxonomic [11, 83-87] and functional [13, 16, 22] profiles of IBD patients compared to healthy controls, especially in Crohn's disease. Additionally, clinical administration of antibiotics shows promise for reducing the intestinal inflammation associated with IBD [88, 89]. The longitudinal biomarkers we identified are promising new candidates to investigate in the context of human disease onset and progression.

Our analyses identified several modules that implicate a pathogenic effect by the DNR microbiome. For example, DNR mouse microbiomes increase in the abundance of adhesion protein transport modules (M00330) in association with disease, which may help pathobiotic members of the microbiome associate with and metabolize intestinal mucosa [90]. Correspondingly, keratan (M00079) and

dermatan (M00076) sulfate degradation pathways increase in abundance as disease progresses. Keratan sulfate and dermatan sulfate are glycosaminoglycans (GAGs) that are integral to intestinal mucosa and regulate the permeability of the gut epithelium. These sulfated GAGs are depleted in IBD patients [91], and their metabolism by intestinal bacteria, including *Bacteroides thetaiotaomicron*, contributes to intestinal colonization [92, 93]. Furthermore, Crohn's metagenomes exhibit an increase in GAG degradation pathways [94]. DNR guts also have elevated levels of Type III and Type IV secretion systems, which pathogenic organisms leverage to successfully invade the gut microbiome and induce preferable ecological conditions within the gut [68, 69]. Curiously, type III secretion abundance shows the opposite effect before immune activation (weeks 4-7), perhaps because of broad shifts in community composition after week 7 or alternatively due to microbes with type III secretion systems invading the LP and becoming less abundant in stool over time. Finally, we observe an increase in modules associated with the biosynthesis of isoprenoids, which have been linked to the stimulation of the mammalian immune system [95]. Together these DNR-associated pathways support a pathogenic role of gut microbes in IBD development. Future studies that seek to determine the existence of a microbiome-mediated etiology for IBD should consider these potential mechanisms of disease activation.

Our identification of pathways that change in association with IBD development generates many novel hypotheses about the mechanisms through which gut microbes contribute or respond to disease development. Future studies can explicitly test these hypotheses to discern the cause and effect relationship between the gut microbiome and inflammatory bowel disease. Several KEGG modules with different abundance dynamics in DNR versus WT mice appear to be associated with the microbiome's acclimation to the disease environment. For example, we observe increases in two-component systems (M00511, M00482) that may contribute to a cell's ability to manage the elevated oxidative stress that exists during active IBD [96]. We also observe increases in pathways associated with cellular chemotaxis (M00515, M00507). This result is consistent with observations of increased cell motility pathways in the gut microbiomes of TRUC mice suffering active colitis using imputation from 16S data

[22]. This result also aligns with prior work that implicated toll-like receptor recognition of flagellar bacterial antigens in the development of intestinal inflammation [97, 98]. Based on these observations, we speculate that chemotaxis pathways help microbiota scavenge the metabolic resources required to survive inside of an inflamed gut or invade the host given that intestinal permeability frequently increases during IBD flare-ups [99].

We also observe several biosynthetic modules that increase in association with IBD development. For example, modules related to the biosynthesis of uridine monophosphate, leucine, proline, and ammonia change in association with disease. These results may suggest that the metabolic preferences and needs of the organisms that comprise the microbiome change as disease develops. Alternatively, it may be that more T cells are entering the gut, becoming activated, and consequently consuming the local resources, which in turn results in bacteria activating biosynthetic pathways to survive and compete. Our finding that pathways associated with ammonia production (M00531) increase in DNR mice is noteworthy because prior studies have found that IBD associates with a lower pH in the intestinal lumen [100], and the production of ammonia by bacteria may serve to buffer such pH changes. Additionally, these pathways are utilized when bacteria metabolize proteins, amino acids, and urea, and the increase in this pathway may indicate a preferential utilization of these substrates by the microbiome or, as above, immune cells, during disease.

Furthermore, we observe increases in modules associated with choline metabolism, specifically betaine and phosphatidylcholine biosynthesis. Recent work has connected the gut microbiome's production of these metabolites to increased cardiovascular disease risk [101]. Our finding is important because a growing number of studies indicate that IBD patients have an increased risk of developing cardiovascular disease, especially during flare-ups [102, 103]. The mechanisms underlying this increased risk are not well resolved, but may relate to a proposed explanation for the increased cardiovascular disease risk observed in HIV-infected patients [104, 105]. Under this model, changes in the relative proportion of protective to pathobiotic gut microbiota, especially those capable of

translocating across the gut epithelium, activate a chronic systemic inflammation that increases cardiovascular disease risk. It is thus tempting to speculate that, based on our observations in these mouse models of disease, IBD and perhaps HIV-associated changes in microbial metabolism of choline contribute to or at least indicate this increased risk of cardiovascular disease.

Another intriguing hypothesis emerges from our observation that heme transport genes are elevated in DNR mice as IBD develops. Bacteria use this module to scavenge iron from the environment. Iron is a crucial component for many cellular processes, but gut microbes seldom have access to free iron and instead sequester it from host sources, such as heme [106, 107]. Heme concentrations may be increased in IBD, as a common feature of the disease is intestinal bleeding [108]. Hence, we hypothesize that gut microbes that can take advantage of this heme may flourish in DNR mice. It is intriguing to further speculate that microbial sequestration of heme contributes to IBD (e.g., through signaling to the immune system) or to iron deficiency in IBD patients [109].

One surprising discovery was an increase in pathways associated with the production of benzoate (M00538) in DNR mice. Benzoate is a carboxylic acid produced by microbial degradation of dietary aromatic compounds and is a precursor of hippurate biosynthesis in mammals [110]. Prior work suggested that hippurate may be a useful diagnostic of Crohn's disease given that it is found at significantly lower levels in the urine of patients [110] and that the gut microbiome's production of benzoate is responsible for these differences in urinary hippurate [111]. Our results are inconsistent with this prior work in such that they indicate that intestinal benzoate biosynthesis is higher in sick animals. This difference may be due to variation in the host species being investigated, including how benzoate is subsequently metabolized in the gut or by the host. Alternatively, the potential of the DNR microbiome to make excess hippurate may not be realized given that we performed DNA sequencing. Future mechanistic studies could measure benzoate and hippurate and quantify the benzoate proteins at the RNA or protein level in DNR versus WT mice.

Most of the taxonomic and functional IBD biomarkers we identified are increasingly abundant in DNR mice throughout the disease process. But three modules show the opposite trajectory and decrease in abundance over time in DNR relative to WT mice: melatonin biosynthesis (M00037), lysine biosynthesis (M00031), and lipooligosaccharide transport (M00252). Melatonin has a dual effect on the immune system, acting in a stimulatory manner in early infection, and in an immunomodulatory manner in cases of prolonged inflammation [112]. The effects of melatonin produced by gut commensals have not been studied as extensively as those of endogenous melatonin. Traditionally, melatonin acts as a potent antioxidant, although additional quorum signaling functions in bacteria have been recently reported [113]. The reduction in melatonin biosynthesis capacity observed in the DNR mice could be caused by the expansion of species that can tolerate a highly oxidative environment [114] or microbes that utilize other strategies for neutralizing reactive oxygen species. Without metabolite data, it is not possible to definitively say that the final concentrations of melatonin are reduced in the disease state, since the decrease can be offset by host production. With respect to lysine biosynthesis, this module is also depleted in human IBD microbiomes [115], indicating that there may exist similar mechanisms of interaction between disease context and the gut microbiome across species. Future work should empirically test the potential role of these microbiome functions on the development of IBD, especially in individuals that are genetically susceptible for the disease.

Lipooligosaccharide transport is the only module to show significant differences in abundance trajectories both pre- and post-activation. Intriguingly, it is consistently lower in DNR versus WT mice throughout our study with the largest difference during weeks 4-7, prior to immune activation and disease symptoms. This finding initially seems surprising, because lipooligosaccharides are the major glycolipids that are produced by mucosal Gram-negative bacteria and are known to have proinflammatory effects [116]. However, the two genes (NodI and NodJ) in the lipooligosaccharide transport system are present across diverse prokaryotes, and the substrates of this two-component ABC transporter are not characterized beyond lipo-chitin oligosaccharide export in rhizobial bacteria [117, 118]. Determining what this system transports in the mammalian gut and how its function changes

in IBD is an exciting future direction. Regardless of mechanism, the consistent and pre-symptomatic depletion of lipooligosaccharide transport genes in DNR mice make this module a promising candidate biomarker for predicting and diagnosing IBD.

We relied on a mouse model to quantify the longitudinal interaction between the gut microbiome and disease because the extensive inter-individual variation in human genetics, lifestyle, microbiome composition, and disease status and severity can complicate study design, analysis, and interpretation. We used the DNR mouse model because it is relevant to our understanding of the mucosal immunological dysregulation that occurs during human IBD and, consequently, its interaction with the gut microbiome. Indeed, we observe immune activation in the blood of the DNR mice that is consistent with what has been observed in human IBD [119]. The phenotype observed in DNR mice is akin to severe Crohn's disease with relatively substantial immunological activation and weight loss by week 12. Interpretations of the microbiome-disease interaction in this model should be considerate of this relatively severe disease status. Alternative mouse lines may be better models for other forms of IBD. Another consideration is that we found some baseline differences in microbiome protein abundances in DNR mice at weaning that may be specific to this genetic model of IBD. Ultimately, comparisons between our results and those obtained by the integrated Human Microbiome Project (iHMP) [120] which is longitudinally evaluating the microbiome and immune status of IBD patients, will clarify the relevance of the findings produced by the DNR model to human populations. Additionally, future research should use this model and build upon our findings to clarify how TGF-β induced differentiation and function of T cells interacts with the taxonomic structure and function of the gut microbiome.

Overall, our results indicate that the development of IBD is associated with corresponding changes in the operation of the gut microbiome. Microbial taxa and KEGG module abundances vary over time and in association with immune activation. Furthermore, our results suggest that the gut microbiome may contribute to disease by activating inflammation through metabolism of mucosa and by expressing proinflammatory and downregulating anti-inflammatory metabolites. Because our study

485 relied on the imputation of microbiome function from DNA sequences, we cannot definitively conclude
486 that the observed differences in the microbiome's functional profiles manifest as differences in the
487 metabolites produced by the microbiome. Future research that applies direct measurements of
488 microbiome function should be used to validate and expand the results presented here. Regardless,
489 our results hold promise for our understanding of microbiome-mediated IBD disease mechanisms and
490 the potential of using microbiome sequencing of patient stool to classify and potentially even predict
491 disease.
492
493 **Methods**
494 *Growth of mice and microbiome sampling*
495 We bred two cohorts of DNR and WT littermate control animals in the Gladstone Institutes mouse
496 facility as follows. CD4-dnTβRII (DNR) animals were crossed to RAG1-/- background to eliminate the T
497 cell mediated IBD, and were transferred from Yale University to Gladstone Institutes in 2010. To initiate
498 experiments described in this study, DNR-RAG1-/- males were bred with C57BL/6N female animals,
499 and DNR-RAG1-/+ progeny were again crossed to C57BL/6N females to generate a combination of
500 RAG1-/+ and RAG1+/+ DNR and WT age-matched littermate controls. Animals were given regular
501 chow consisting of irradiated PicoLab Rodent Diet 20 (LabDiet). Only female animals were used in this
502 study. Four co-housed WT-RAG1+/+ and five co-housed DNR-RAG1+/+ littermates were followed
503 longitudinally for 15 weeks and fresh fecal samples were collected weekly and stored at -30° C until
504 they were subject to microbiome processing. All mice from both cohorts were weighed weekly.
505 All animal experiments were conducted in accordance with guidelines set by the Institutional Animal
506 Care and Use Committee of the University of California, San Francisco.
507
508 *Immune sampling*
509 Tail vein blood samples were collected weekly from a parallel cohort of "bleeder" mice to quantify how
510 their immune status changes over time (n=6 WT, n=6 DNR). These are distinct individuals from the
511 "pooper" mice subjected to stool metagenomics (same colony and time period), in order to prevent

repeated tail vein blood sampling from affecting the health or microbiota of the cohort of pooper mice. Specifically, ~100 µl (2-3 drops) of blood from tail vein was added to 30 µl of 1x heparin (500 units/ml). 500 µl of 1x ACK lysis buffer (Lonza) was added directly to the cells and incubated at room temperature for 2–3 min. Cells were centrifuged at 4000 rpm for 5 min. The top layer was aspirated and another 500 ml of 1x ACK lysis buffer was added followed by centrifugation. Cells were resuspended in FACS buffer (PBS + 0.5% FBS) and after blocking Fc receptors with anti-CD16/CD32, single-cell suspensions were incubated with FITC CD4 (GK1.5), PE CD62L (MEL14), PerCP-Cy5.5 CD8a (53.6.72), and APC CD44 (IM7) mouse antibodies for 30 min at 4°C. Stained cells were washed and acquired on an Accuri C6 cytometer (BD). Blood lymphocytes were gated on CD4+ or CD8+ fractions and percentage of activated/memory (CD44hi) among CD4+ and CD8+ T cells was determined using FlowJo software (Tree Star Inc.). This cohort was separated from those subjected to microbiome sampling to eliminate the effect that repeated bloodletting might have on the microbiome. At 15 weeks of age, two WT and three DNR from the non-bleeding animals were euthanized. Spleen and mesenteric lymph nodes were then processed into single-cell suspension and subjected to ACK lysis and cell surface staining as described for PBMCs. The status of T cell activation was quantified and found to highly correlate with the blood immune status of their “bleeder” littermates (Additional Data File 2: Figure S2 and Additional Data File 3: Table S1).

Metagenome sequencing and analysis

QIAamp DNA Stool Mini Kits (QIAGEN, Valencia, CA, USA) were used to extract DNA from stool samples collected from weeks 4, 6, 8, 10, and 12. Samples were incubated in an elevated water bath temperature of 95 C to increase the lysis of bacterial cells, as per manufacturer instructions. The MoBio PowerFecal DNA isolation kit (MOBIO, Carlsbad, CA USA) was used as per manufacturer instructions to process stool samples collected at weeks 5 and 13. Kit type was adjusted for in statistical modeling to account for any potential differences in extraction bias between the two methods.

Purified DNA was prepared for shotgun metagenomic sequencing using the Nextera XT library preparation method (ILLUMINA, San Diego, CA USA). Libraries were quality assessed using qPCR and a Bioanalyzer (Agilent Technologies, Palo Alto, CA USA) and subsequently sequenced using an

Illumina HiSeq 2000. This produced an average of 74,427,303 100-bp paired-end sequences per sample. Metagenomic reads were quality controlled using the standard operating procedure defined by the Human Microbiome Project Consortium [121] as implemented in shotcleaner [122]. Briefly, reads were quality trimmed using prinseq [123] and mapped against the mouse reference genome sequence (GRCm38) using bmtagger [124]. Exact duplicate reads were collapsed and the subsequent high-quality data was subject to taxonomic and functional annotation. Functional annotation of metagenomes was conducted using ShotMAP as described in [16] with, Prodigal [125] to call genes and RAPsearch2 [126] to identify metagenomic homologs of the KEGG database (downloaded Feb. 2015). Reads mapping to mammalian sequences in the KEGG database were discarded and the subsequent data was used to quantify the abundance of each KEGG Orthology Group (KO) using the RPKG abundance statistic [127]. Metagenomes were taxonomically annotated using MIDAS as described in [72].

Statistical analyses and modeling

The functional and taxonomic similarity between metagenomic samples was assessed using non-metric multi-dimensional scaling (NMDS) as implemented through the nmds function in the labdsv R package [128]. Ordinations were visualized using the ordiplot function in the vegan R package [129]. For the functional similarity analysis, the vegdist function from the vegan R package quantified the Bray-Curtis dissimilarity based on KEGG module abundances. The taxonomic analysis used the generalized Unifrac [130] distances (alpha=0.5), which were obtained by using the taxonomic tree from the Living Tree Project [131] and matching the genus and species component of tree leaf labels to the corresponding components of the MIDAS species labels in our data. Assessment of the significance of the clustering of samples in these ordination plots was conducted using PERMANOVA as implemented by the adonis function in R.

The compound Poisson generalized linear mixed effects model implemented in the cplm package in R [132] was used to find KEGG modules with a significantly different time trend between groups while controlling for static differences between the lines and DNA extraction procedure

(QIAGEN versus MOBIO). Random intercepts and slopes for both subjects and contributing KOs were used to capture variation between subjects and between genes while focusing on the large scale shifts over the whole collection of abundance profiles contributing to a module. As described more thoroughly in Supplementary Text S1, the general computational procedure consisted of subsetting the data to each module's relevant KO abundances and fitting a full model that described the RPKG abundance as a function of time, group, time by group interaction, sequencing kit and random effects of each KO and individual. We then use two reduced models, dropping first the interaction term and then the group term, to obtain p-values via likelihood ratio tests. This is one of the recommended significance testing approaches for mixed models since it avoids using approximations for the residual degrees of freedom that would be necessary to test significance via the t-statistic [66]. To limit the number of modules tested, the input data were run through the MinPath algorithm [133] to select a parsimonious set of modules based on the KOs present. The union of all samples' individual parsimonious sets was used as the final set of tested modules. The approach of testing the dynamics of an entire module by fitting a single GLMM to a set of multiple genes' temporal abundances is modeled on the TcGSA method of Hejblum et al. [134], with the modification of using a different response distribution (the Tweedie compound Poisson). Significant modules were selected at the 0.05 FDR threshold after controlling for multiple testing via the Benjamini-Hochberg procedure (B-H). Species time trend differences were tested with the same approach, minus the grouping of multiple trajectories. Additional details of our modeling approach can be found in Additional Data File 5: Text S1. All of the code used in this analysis is available at the following URL:

https://github.com/slyalina/Mouse_IBD_2017_paper_supporting_code.

To differentiate functional changes occurring prior to immune activation, we fit a second hinge regression to the abundances of modules that were found to have a significant time by group interaction in the main GLMM analysis. This second regression placed a break point at week 7, which represents the point at which immune activation initiated (Fig. 1). This allowed for two sets of slopes (before disease onset and after) and two sets of time by group interactions (representing deviations of DNR slopes from WT before-onset and after-onset slopes).

Alterations in the species trajectory curves were additionally tested with an alternate method aimed at highlighting differences in shape rather than slope. This method was an implementation of the FPCA-based difference in goodness-of-fit approach described previously in [135]. The permutation-based p-values from this analysis were B-H corrected and species passing the 0.05 FDR threshold were retained.

To test the hypothesis that the distribution of all modules' between-KO/within-KO dispersion decomposition statistic is significantly different from random when grouping functional trajectories (KOs into modules) but is not significantly different from random when grouping taxonomic trajectories (species into genera) we used the DISCO [71] non-parametric test to obtain the real distributions of the test statistic in the two scenarios, as well the simulated null distributions that arise when generating random groupings of KOs and species. We then performed a Kolmogorov-Smirnov test to compare the true distributions with their simulated counterparts.

Accession numbers

Metagenomic sequences are available through GenBank accessions SAMN06921515 - SAMN06921563.

Acknowledgements

We are grateful to Stephen Nayfach for helpful discussions. We also appreciate the services provided by the Gladstone genomics core, UCSF sequencing core, and the Center for Genome Research and Biocomputing.

Competing Interests

The authors declare that they have no competing interests.

Funding

This project was supported by NIAID R21 grant #AI108953 and the Gladstone Institutes.

620

621

622 **Figure Legends**

623

624 **Figure 1. IBD development correlates with peripheral T cell activation in DNR mice**. (**A**) Animal
625 weight over time. N= 7 WT and 8 DNR mice. (**B**) Percent of activated CD4 T cells among peripheral
626 blood mono-nuclear cells (PBMCs). (**C**) Percent of activated CD8 T cells among PBMCs. (**B-C**) N=6
627 WT and 6 DNR mice.

628

629 **Figure 2. The taxonomic and functional diversity of the gut microbiome associates with IBD**
630 **development.** (A) NMDS ordination plots of the functional (left) and taxonomic (right) beta-diversity of
631 samples from each line illustrate the significant divergence in beta-diversity between lines over time.
632 Functional beta-diversity was measured as the Bray-Curtis dissimilarity based on KEGG module
633 abundances, while taxonomic beta-diversity is the UniFrac distance of taxa detected in metagenomes.
634 (B) The longitudinal variation of samples along selected NMDS dimensions similarly reveals how DNR
635 and WT lines significantly diverge over time both in terms of their functional (left) and taxonomic (right)
636 beta-diversity. Plotted is the smoothed LOESS trajectory of samples from each line over time, where
637 grey areas represent 95% confidence intervals.

638

639 **Figure 3. Summary of GLMM results of 29 modules with significant time by group interaction.**
640 (A) The quantity plotted is the predicted marginal mean (PMM) of the slope coefficients. Significance
641 testing was done by comparing goodness of fit of full and reduced GLMM specifications, and the full
642 model was used to produce the PMM estimates shown here. This quantity was primarily calculated to
643 get a succinct summary of the direction of temporal change, and does not always coincide with the
644 interaction coefficient that is the focus of the main analysis. The estimates were obtained by running the
645 lstrends function from the lsmeans R package [136] (B) The underlying KO abundance trajectories of a
646 significant module (M00031: Lysine biosynthesis) that decreases in DNR mice and increases in WT
647 mice over time, as evidenced by a negative and positive model slope, respectively. (C) A similar plot as
648 in (B), except that this significant module (M00330: Adhesin transport) significantly increases in

649 abundance over time in DNR mice, while it does not change in abundance in WT mice. For both (B)
650 and (C) the shaded ribbons represent LOESS confidence bounds.
651
652 **Figure 4. Modules with significantly differing slopes between groups show primarily post-**
653 **disease-onset differences when analyzed with a segmented GLMM**. For each cohort, the
654 segmented GLMM estimates two separate WT slopes (pre-week7 and post-week7) and two deviations
655 from those slopes, which represent the time by group interaction that measures how DNR slopes differ
656 from WT slopes. Plotted here are the estimates of these deviations, with asterisks marking coefficients
657 that were significantly non-zero with B-H corrected p-value of <0.2
658
659 **Figure 5. Species that showed significantly different trajectory shapes between DNR and WT**
660 **groups.** These results are based on an FPCA-based goodness-of-fit comparison test that identified 7
661 species that were different at B-H corrected p-value < 0.05.
662
663

664 **Table Legends**

665

666 **Table 1** - Species with significantly different trajectory shapes in the FPCA-based goodness-of-fit

667 comparisons.

668

References

1. Huttenhower C, Kostic AD, Xavier RJ. 2014. Inflammatory bowel disease as a model for translating the microbiome. Immunity 40:843-54.
2. Wlodarska M, Kostic AD, Xavier RJ. 2015. An integrative view of microbiome-host interactions in inflammatory bowel diseases. Cell Host Microbe 17:577-91.
3. Ayres JS. 2016. Cooperative Microbial Tolerance Behaviors in Host-Microbiota Mutualism. Cell 165:1323-31.
4. Kostic AD, Xavier RJ, Gevers D. 2014. The microbiome in inflammatory bowel disease: current status and the future ahead. Gastroenterology 146:1489-99.
5. Lees CW, Barrett JC, Parkes M, Satsangi J. 2011. New IBD genetics: common pathways with other diseases. Gut 60:1739-53.
6. Wirtz S, Neurath MF. 2007. Mouse models of inflammatory bowel disease. Adv Drug Deliv Rev 59:1073-83.
7. Goodrich JK, Waters JL, Poole AC, Sutter JL, Koren O, Blekhman R, Beaumont M, Van Treuren W, Knight R, Bell JT, Spector TD, Clark AG, Ley RE. 2014. Human genetics shape the gut microbiome. Cell 159:789-99.
8. Neuman MG, Nanau RM. 2012. Inflammatory bowel disease: role of diet, microbiota, life style. Transl Res 160:29-44.
9. Bilski J, Mazur-Bialy A, Brzozowski B, Magierowski M, Zahradnik-Bilska J, Wojcik D, Magierowska K, Kwiecien S, Mach T, Brzozowski T. 2016. Can exercise affect the course of inflammatory bowel disease? Experimental and clinical evidence. Pharmacol Rep 68:827-36.
10. Sartor RB, Mazmanian SK. 2012. Intestinal Microbes in Inflammatory Bowel Diseases. Am J Gastroenterol Suppl 1:15-21.
11. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, Nielsen T, Pons N, Levenez F, Yamada T, Mende DR, Li J, Xu J, Li S, Li D, Cao J, Wang B, Liang H, Zheng H, Xie Y, Tap J, Lepage P, Bertalan M, Batto JM, Hansen T, Le Paslier D, Linneberg A, Nielsen HB, Pelletier E, Renault P, Sicheritz-Ponten T, Turner K, Zhu H, Yu C, Li S, Jian M, Zhou Y, Li Y, Zhang X, Li S, Qin N, Yang H, Wang J, Brunak S, Dore J, Guarner F, Kristiansen K, Pedersen O, Parkhill J, Weissenbach J, et al. 2010. A human gut microbial gene catalogue established by metagenomic sequencing. Nature 464:59-65.
12. Nielsen HB, Almeida M, Juncker AS, Rasmussen S, Li J, Sunagawa S, Plichta DR, Gautier L, Pedersen AG, Le Chatelier E, Pelletier E, Bonde I, Nielsen T, Manichanh C, Arumugam M, Batto JM, Quintanilha Dos Santos MB, Blom N, Borruel N, Burgdorf KS, Boumezbeur F, Casellas F, Dore J, Dworzynski P, Guarner F, Hansen T, Hildebrand F, Kaas RS, Kennedy S, Kristiansen K, Kultima JR, Leonard P, Levenez F, Lund O, Moumen B, Le Paslier D, Pons N, Pedersen O, Prifti E, Qin J, Raes J, Sorensen S, Tap J, Tims S, Ussery DW, Yamada T, Meta HITC, Renault P, Sicheritz-Ponten T, Bork P, et al. 2014. Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. Nat Biotechnol 32:822-8.
13. Morgan XC, Tickle TL, Sokol H, Gevers D, Devaney KL, Ward DV, Reyes JA, Shah SA, LeLeiko N, Snapper SB, Bousvaros A, Korzenik J, Sands BE, Xavier RJ, Huttenhower C. 2012. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. Genome Biol 13:R79.

14. Ott SJ, Musfeldt M, Wenderoth DF, Hampe J, Brant O, Folsch UR, Timmis KN, Schreiber S. 2004. Reduction in diversity of the colonic mucosa associated bacterial microflora in patients with active inflammatory bowel disease. Gut 53:685-93.
15. Dubinsky M, Braun J. 2015. Diagnostic and Prognostic Microbial Biomarkers in Inflammatory Bowel Diseases. Gastroenterology 149:1265-1274 e3.
16. Nayfach S, Bradley PH, Wyman SK, Laurent TJ, Williams A, Eisen JA, Pollard KS, Sharpton TJ. 2015. Automated and Accurate Estimation of Gene Family Abundance from Shotgun Metagenomes. PLoS Comput Biol 11:e1004573.
17. Atarashi K, Tanoue T, Shima T, Imaoka A, Kuwahara T, Momose Y, Cheng G, Yamasaki S, Saito T, Ohba Y, Taniguchi T, Takeda K, Hori S, Ivanov, II, Umesaki Y, Itoh K, Honda K. 2011. Induction of colonic regulatory T cells by indigenous Clostridium species. Science 331:337-41.
18. Honda K, Littman DR. 2012. The microbiome in infectious disease and inflammation. Annu Rev Immunol 30:759-95.
19. Chu H, Khosravi A, Kusumawardhani IP, Kwon AH, Vasconcelos AC, Cunha LD, Mayer AE, Shen Y, Wu WL, Kambal A, Targan SR, Xavier RJ, Ernst PB, Green DR, McGovern DP, Virgin HW, Mazmanian SK. 2016. Gene-microbiota interactions contribute to the pathogenesis of inflammatory bowel disease. Science 352:1116-20.
20. David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, Wolfe BE, Ling AV, Devlin AS, Varma Y, Fischbach MA, Biddinger SB, Dutton RJ, Turnbaugh PJ. 2014. Diet rapidly and reproducibly alters the human gut microbiome. Nature 505:559-63.
21. Carmody RN, Gerber GK, Luevano JM, Jr., Gatti DM, Somes L, Svenson KL, Turnbaugh PJ. 2015. Diet dominates host genotype in shaping the murine gut microbiota. Cell Host Microbe 17:72-84.
22. Rooks MG, Veiga P, Wardwell-Scott LH, Tickle T, Segata N, Michaud M, Gallini CA, Beal C, van Hylckama-Vlieg JE, Ballal SA, Morgan XC, Glickman JN, Gevers D, Huttenhower C, Garrett WS. 2014. Gut microbiome composition and function in experimental colitis during active disease and treatment-induced remission. ISME J 8:1403-17.
23. Strober W, Fuss IJ, Blumberg RS. 2002. The immunology of mucosal models of inflammation. Annu Rev Immunol 20:495-549.
24. Bouma G, Strober W. 2003. The immunological and genetic basis of inflammatory bowel disease. Nat Rev Immunol 3:521-33.
25. Hoffmann JC, Pawlowski NN, Kuhl AA, Hohne W, Zeitz M. 2002. Animal models of inflammatory bowel disease: an overview. Pathobiology 70:121-30.
26. Mizoguchi A, Mizoguchi E, Bhan AK. 2003. Immune networks in animal models of inflammatory bowel disease. Inflamm Bowel Dis 9:246-59.
27. Elson CO, Cong Y, McCracken VJ, Dimmitt RA, Lorenz RG, Weaver CT. 2005. Experimental models of inflammatory bowel disease reveal innate, adaptive, and regulatory mechanisms of host dialogue with the microbiota. Immunol Rev 206:260-76.
28. Valatas V, Vakas M, Kolios G. 2013. The value of experimental models of colitis in predicting efficacy of biological therapies for inflammatory bowel diseases. Am J Physiol Gastrointest Liver Physiol 305:G763-85.
29. Valatas V, Bamias G, Kolios G. 2015. Experimental colitis models: Insights into the pathogenesis of inflammatory bowel disease and translational issues. Eur J Pharmacol 759:253-64.
30. Gorelik L, Flavell RA. 2000. Abrogation of TGFbeta signaling in T cells leads to spontaneous T cell differentiation and autoimmune disease. Immunity 12:171-81.

31. Yamagiwa S, Gray JD, Hashimoto S, Horwitz DA. 2001. A role for TGF-beta in the generation and expansion of CD4+CD25+ regulatory T cells from human peripheral blood. J Immunol 166:7282-9.
32. Chen W, Jin W, Hardegen N, Lei KJ, Li L, Marinos N, McGrady G, Wahl SM. 2003. Conversion of peripheral CD4+CD25- naive T cells to CD4+CD25+ regulatory T cells by TGF-beta induction of transcription factor Foxp3. J Exp Med 198:1875-86.
33. Schlenner SM, Weigmann B, Ruan Q, Chen Y, von Boehmer H. 2012. Smad3 binding to the foxp3 enhancer is dispensable for the development of regulatory T cells with the exception of the gut. J Exp Med 209:1529-35.
34. Selvaraj RK, Geiger TL. 2007. A kinetic and dynamic analysis of Foxp3 induced in T cells by TGF-beta. J Immunol 178:7667-77.
35. Fantini MC, Becker C, Monteleone G, Pallone F, Galle PR, Neurath MF. 2004. Cutting edge: TGF-beta induces a regulatory phenotype in CD4+CD25- T cells through Foxp3 induction and down-regulation of Smad7. J Immunol 172:5149-53.
36. Huber S, Schramm C, Lehr HA, Mann A, Schmitt S, Becker C, Protschka M, Galle PR, Neurath MF, Blessing M. 2004. Cutting edge: TGF-beta signaling is required for the in vivo expansion and immunosuppressive capacity of regulatory CD4+CD25+ T cells. J Immunol 173:6526-31.
37. Marie JC, Letterio JJ, Gavin M, Rudensky AY. 2005. TGF-beta1 maintains suppressor function and Foxp3 expression in CD4+CD25+ regulatory T cells. J Exp Med 201:1061-7.
38. Battaglia A, Buzzonetti A, Baranello C, Fanelli M, Fossati M, Catzola V, Scambia G, Fattorossi A. 2013. Interleukin-21 (IL-21) synergizes with IL-2 to enhance T-cell receptor-induced human T-cell proliferation and counteracts IL-2/transforming growth factor-beta-induced regulatory T-cell development. Immunology 139:109-20.
39. Molinero LL, Miller ML, Evaristo C, Alegre ML. 2011. High TCR stimuli prevent induced regulatory T cell differentiation in a NF-kappaB-dependent manner. J Immunol 186:4609-17.
40. Bettelli E, Carrier Y, Gao W, Korn T, Strom TB, Oukka M, Weiner HL, Kuchroo VK. 2006. Reciprocal developmental pathways for the generation of pathogenic effector TH17 and regulatory T cells. Nature 441:235-8.
41. Veldhoen M, Stockinger B. 2006. TGFbeta1, a "Jack of all trades": the link with pro-inflammatory IL-17-producing T cells. Trends Immunol 27:358-61.
42. Li MO, Wan YY, Flavell RA. 2007. T cell-produced transforming growth factor-beta1 controls T cell tolerance and regulates Th1- and Th17-cell differentiation. Immunity 26:579-91.
43. Gutcher I, Donkor MK, Ma Q, Rudensky AY, Flavell RA, Li MO. 2011. Autocrine transforming growth factor-beta1 promotes in vivo Th17 cell differentiation. Immunity 34:396-408.
44. Manel N, Unutmaz D, Littman DR. 2008. The differentiation of human T(H)-17 cells requires transforming growth factor-beta and induction of the nuclear receptor RORgammat. Nat Immunol 9:641-9.
45. Yang L, Anderson DE, Baecher-Allan C, Hastings WD, Bettelli E, Oukka M, Kuchroo VK, Hafler DA. 2008. IL-21 and TGF-beta are required for differentiation of human T(H)17 cells. Nature 454:350-2.
46. Biancheri P, Giuffrida P, Docena GH, MacDonald TT, Corazza GR, Di Sabatino A. 2014. The role of transforming growth factor (TGF)-beta in modulating the immune response and fibrogenesis in the gut. Cytokine Growth Factor Rev 25:45-55.

47. Di Giacinto C, Marinaro M, Sanchez M, Strober W, Boirivant M. 2005. Probiotics ameliorate recurrent Th1-mediated murine colitis by inducing IL-10 and IL-10-dependent TGF-beta-bearing regulatory cells. J Immunol 174:3237-46.
48. Feagins LA. 2010. Role of transforming growth factor-beta in inflammatory bowel disease and colitis-associated colon cancer. Inflamm Bowel Dis 16:1963-8.
49. Fuss IJ, Boirivant M, Lacy B, Strober W. 2002. The interrelated roles of TGF-beta and IL-10 in the regulation of experimental colitis. J Immunol 168:900-8.
50. Harrison OJ, Powrie FM. 2013. Regulatory T cells and immune tolerance in the intestine. Cold Spring Harb Perspect Biol 5.
51. Izcue A, Coombes JL, Powrie F. 2009. Regulatory lymphocytes and intestinal inflammation. Annu Rev Immunol 27:313-38.
52. Jarry A, Bossard C, Sarrabayrouse G, Mosnier JF, Laboisse CL. 2011. Loss of interleukin-10 or transforming growth factor beta signaling in the human colon initiates a T-helper 1 response via distinct pathways. Gastroenterology 141:1887-96 e1-2.
53. Kitani A, Fuss I, Nakamura K, Kumaki F, Usui T, Strober W. 2003. Transforming growth factor (TGF)-beta1-producing regulatory T cells induce Smad-mediated interleukin 10 secretion that facilitates coordinated immunoregulatory activity and amelioration of TGF-beta1-mediated fibrosis. J Exp Med 198:1179-88.
54. Weiner HL. 2001. Induction and mechanism of action of transforming growth factor-beta-secreting Th3 regulatory cells. Immunol Rev 182:207-14.
55. Franke A, McGovern DP, Barrett JC, Wang K, Radford-Smith GL, Ahmad T, Lees CW, Balschun T, Lee J, Roberts R, Anderson CA, Bis JC, Bumpstead S, Ellinghaus D, Festen EM, Georges M, Green T, Haritunians T, Jostins L, Latiano A, Mathew CG, Montgomery GW, Prescott NJ, Raychaudhuri S, Rotter JI, Schumm P, Sharma Y, Simms LA, Taylor KD, Whiteman D, Wijmenga C, Baldassano RN, Barclay M, Bayless TM, Brand S, Buning C, Cohen A, Colombel JF, Cottone M, Stronati L, Denson T, De Vos M, D'Inca R, Dubinsky M, Edwards C, Florin T, Franchimont D, Gearry R, Glas J, Van Gossum A, et al. 2010. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. Nat Genet 42:1118-25.
56. Glocker EO, Kotlarz D, Boztug K, Gertz EM, Schaffer AA, Noyan F, Perro M, Diestelhorst J, Allroth A, Murugan D, Hatscher N, Pfeifer D, Sykora KW, Sauer M, Kreipe H, Lacher M, Nustede R, Woellner C, Baumann U, Salzer U, Koletzko S, Shah N, Segal AW, Sauerbrey A, Buderus S, Snapper SB, Grimbacher B, Klein C. 2009. Inflammatory bowel disease and mutations affecting the interleukin-10 receptor. The New England journal of medicine 361:2033-45.
57. McGovern DP, Gardet A, Torkvist L, Goyette P, Essers J, Taylor KD, Neale BM, Ong RT, Lagace C, Li C, Green T, Stevens CR, Beauchamp C, Fleshner PR, Carlson M, D'Amato M, Halfvarson J, Hibberd ML, Lordal M, Padyukov L, Andriulli A, Colombo E, Latiano A, Palmieri O, Bernard EJ, Deslandres C, Hommes DW, de Jong DJ, Stokkers PC, Weersma RK, Sharma Y, Silverberg MS, Cho JH, Wu J, Roeder K, Brant SR, Schumm LP, Duerr RH, Dubinsky MC, Glazer NL, Haritunians T, Ippoliti A, Melmed GY, Siscovick DS, Vasiliauskas EA, Targan SR, Annese V, Wijmenga C, Pettersson S, Rotter JI, et al. 2010. Genome-wide association identifies multiple ulcerative colitis susceptibility loci. Nat Genet 42:332-7.
58. Naviglio S, Arrigo S, Martelossi S, Villanacci V, Tommasini A, Loganes C, Fabretto A, Vignola S, Lonardi S, Ventura A. 2014. Severe inflammatory bowel disease associated with congenital alteration of transforming growth factor beta signaling. J Crohns Colitis 8:770-4.
59. Kang SS, Bloom SM, Norian LA, Geske MJ, Flavell RA, Stappenbeck TS, Allen PM. 2008. An antibiotic-responsive mouse model of fulminant ulcerative colitis. PLoS Med 5:e41.

60. Fowler SA, Ananthakrishnan AN, Gardet A, Stevens CR, Korzenik JR, Sands BE, Daly MJ, Xavier RJ, Yajnik V. 2014. SMAD3 gene variant is a risk factor for recurrent surgery in patients with Crohn's disease. J Crohns Colitis 8:845-51.
61. Cho JH, Brant SR. 2011. Recent insights into the genetics of inflammatory bowel disease. Gastroenterology 140:1704-12.
62. Henderson P, van Limbergen JE, Wilson DC, Satsangi J, Russell RK. 2011. Genetics of childhood-onset inflammatory bowel disease. Inflamm Bowel Dis 17:346-61.
63. Ardizzone S, Bevivino G, Monteleone G. 2016. Mongersen, an oral Smad7 antisense oligonucleotide, in patients with active Crohn's disease. Therap Adv Gastroenterol 9:527-32.
64. Laudisi F, Dinallo V, Di Fusco D, Monteleone G. 2016. Smad7 and its Potential as Therapeutic Target in Inflammatory Bowel Diseases. Curr Drug Metab 17:303-6.
65. Monteleone G, Boirivant M, Pallone F, MacDonald TT. 2008. TGF-beta1 and Smad7 in the regulation of IBD. Mucosal Immunol 1 Suppl 1:S50-3.
66. Bolker BM, Brooks ME, Clark CJ, Geange SW, Poulsen JR, Stevens MH, White JS. 2009. Generalized linear mixed models: a practical guide for ecology and evolution. Trends Ecol Evol 24:127-35.
67. Szekely GJ, Rizzo ML. 2009. Brownian distance covariance. Ann Appl Stat 3:1236-1265.
68. Baumler AJ, Sperandio V. 2016. Interactions between the microbiota and pathogenic bacteria in the gut. Nature 535:85-93.
69. Byndloss MX, Rivera-Chavez F, Tsolis RM, Baumler AJ. 2016. How bacterial pathogens use type III and type IV secretion systems to facilitate their transmission. Curr Opin Microbiol 35:1-7.
70. Mark Welch JL, Utter DR, Rossetti BJ, Mark Welch DB, Eren AM, Borisy GG. 2014. Dynamics of tongue microbial communities with single-nucleotide resolution using oligotyping. Front Microbiol 5:568.
71. Rizzo ML, Szekely GJ. 2010. DISCO analysis: A nonparametric extension of analysis of variance. Ann Appl Stat 4:1034-1055.
72. Nayfach S, Rodriguez-Mueller B, Garud N, Pollard KS. 2016. An integrated metagenomics pipeline for strain profiling reveals novel patterns of bacterial transmission and biogeography. Genome Res 26:1612-1625.
73. Prindiville TP, Sheikh RA, Cohen SH, Tang YJ, Cantrell MC, Silva J, Jr. 2000. Bacteroides fragilis enterotoxin gene sequences in patients with inflammatory bowel disease. Emerg Infect Dis 6:171-4.
74. Rabizadeh S, Rhee KJ, Wu S, Huso D, Gan CM, Golub JE, Wu X, Zhang M, Sears CL. 2007. Enterotoxigenic bacteroides fragilis: a potential instigator of colitis. Inflamm Bowel Dis 13:1475-83.
75. Bloom SM, Bijanki VN, Nava GM, Sun L, Malvin NP, Donermeyer DL, Dunne WM, Jr., Allen PM, Stappenbeck TS. 2011. Commensal Bacteroides species induce colitis in host-genotype-specific fashion in a mouse model of inflammatory bowel disease. Cell Host Microbe 9:390-403.
76. Conley MN, Wong CP, Duyck KM, Hord N, Ho E, Sharpton TJ. 2016. Aging and serum MCP-1 are associated with gut microbiome composition in a murine model. PeerJ 4:e1854.
77. Sartor RB. 2008. Microbial influences in inflammatory bowel diseases. Gastroenterology 134:577-94.
78. Kaur N, Chen CC, Luther J, Kao JY. 2011. Intestinal dysbiosis in inflammatory bowel disease. Gut Microbes 2:211-6.

79. Sellon RK, Tonkonogy S, Schultz M, Dieleman LA, Grenther W, Balish E, Rennick DM, Sartor RB. 1998. Resident enteric bacteria are necessary for development of spontaneous colitis and immune system activation in interleukin-10-deficient mice. Infect Immun 66:5224-31.
80. Garrett WS, Gallini CA, Yatsunenko T, Michaud M, DuBois A, Delaney ML, Punit S, Karlsson M, Bry L, Glickman JN, Gordon JI, Onderdonk AB, Glimcher LH. 2010. Enterobacteriaceae act in concert with the gut microbiota to induce spontaneous and maternally transmitted colitis. Cell Host Microbe 8:292-300.
81. Chichlowski M, Westwood GS, Abraham SN, Hale LP. 2010. Role of mast cells in inflammatory bowel disease and inflammation-associated colorectal neoplasia in IL-10-deficient mice. PLoS One 5:e12220.
82. Waidmann M, Bechtold O, Frick JS, Lehr HA, Schubert S, Dobrindt U, Loeffler J, Bohn E, Autenrieth IB. 2003. Bacteroides vulgatus protects against Escherichia coli-induced colitis in gnotobiotic interleukin-2-deficient mice. Gastroenterology 125:162-77.
83. Frank DN, St Amand AL, Feldman RA, Boedeker EC, Harpaz N, Pace NR. 2007. Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. Proc Natl Acad Sci U S A 104:13780-5.
84. Sokol H, Seksik P, Rigottier-Gois L, Lay C, Lepage P, Podglajen I, Marteau P, Dore J. 2006. Specificities of the fecal microbiota in inflammatory bowel disease. Inflamm Bowel Dis 12:106-11.
85. Peterson DA, Frank DN, Pace NR, Gordon JI. 2008. Metagenomic approaches for defining the pathogenesis of inflammatory bowel diseases. Cell Host Microbe 3:417-27.
86. Sokol H, Seksik P, Furet JP, Firmesse O, Nion-Larmurier I, Beaugerie L, Cosnes J, Corthier G, Marteau P, Dore J. 2009. Low counts of Faecalibacterium prausnitzii in colitis microbiota. Inflamm Bowel Dis 15:1183-9.
87. Pasolli E, Truong DT, Malik F, Waldron L, Segata N. 2016. Machine Learning Meta-analysis of Large Metagenomic Datasets: Tools and Biological Insights. PLoS Comput Biol 12:e1004977.
88. Casellas F, Borruel N, Papo M, Guarner F, Antolin M, Videla S, Malagelada JR. 1998. Antiinflammatory effects of enterically coated amoxicillin-clavulanic acid in active ulcerative colitis. Inflamm Bowel Dis 4:1-5.
89. Rietdijk ST, D'Haens GR. 2013. Recent developments in the treatment of inflammatory bowel disease. J Dig Dis 14:282-7.
90. Gusils C, Morata V, Gonzalez S. 2004. Determination of bacterial adhesion to intestinal mucus. Methods Mol Biol 268:411-5.
91. Murch SH, MacDonald TT, Walker-Smith JA, Levin M, Lionetti P, Klein NJ. 1993. Disruption of sulphated glycosaminoglycans in intestinal inflammation. Lancet 341:711-4.
92. Benjdia A, Martens EC, Gordon JI, Berteau O. 2011. Sulfatases and a radical S-adenosyl-L-methionine (AdoMet) enzyme are key for mucosal foraging and fitness of the prominent human gut symbiont, Bacteroides thetaiotaomicron. J Biol Chem 286:25973-82.
93. Ulmer JE, Vilen EM, Namburi RB, Benjdia A, Beneteau J, Malleron A, Bonnaffe D, Driguez PA, Descroix K, Lassalle G, Le Narvor C, Sandstrom C, Spillmann D, Berteau O. 2014. Characterization of glycosaminoglycan (GAG) sulfatases from the human gut symbiont Bacteroides thetaiotaomicron reveals the first GAG-specific bacterial endosulfatase. J Biol Chem 289:24289-303.
94. Nayfach S, Bradley PH, Wyman SK, Laurent TJ, Williams A, Eisen JA, Pollard KS, Sharpton TJ. 2015. Automated and accurate estimation of gene family abundance from shotgun metagenomes doi:10.1101/022335.

95. Heuston S, Begley M, Gahan CG, Hill C. 2012. Isoprenoid biosynthesis in bacterial pathogens. Microbiology 158:1389-401.
96. Pereira C, Gracio D, Teixeira JP, Magro F. 2015. Oxidative Stress and DNA Damage: Implications in Inflammatory Bowel Disease. Inflamm Bowel Dis 21:2403-17.
97. Steiner TS. 2007. How flagellin and toll-like receptor 5 contribute to enteric infection. Infect Immun 75:545-52.
98. Cullender TC, Chassaing B, Janzon A, Kumar K, Muller CE, Werner JJ, Angenent LT, Bell ME, Hay AG, Peterson DA, Walter J, Vijay-Kumar M, Gewirtz AT, Ley RE. 2013. Innate and adaptive immunity interact to quench microbiome flagellar motility in the gut. Cell Host Microbe 14:571-81.
99. Michielan A, D'Inca R. 2015. Intestinal Permeability in Inflammatory Bowel Disease: Pathogenesis, Clinical Evaluation, and Therapy of Leaky Gut. Mediators Inflamm 2015:628157.
100. Nugent SG, Kumar D, Rampton DS, Evans DF. 2001. Intestinal luminal pH in inflammatory bowel disease: possible determinants and implications for therapy with aminosalicylates and other drugs. Gut 48:571-7.
101. Wang Z, Klipfell E, Bennett BJ, Koeth R, Levison BS, Dugar B, Feldstein AE, Britt EB, Fu X, Chung YM, Wu Y, Schauer P, Smith JD, Allayee H, Tang WH, DiDonato JA, Lusis AJ, Hazen SL. 2011. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. Nature 472:57-63.
102. Schicho R, Marsche G, Storr M. 2015. Cardiovascular complications in inflammatory bowel disease. Curr Drug Targets 16:181-8.
103. Andersen NN, Jess T. 2014. Risk of cardiovascular disease in inflammatory bowel disease. World J Gastrointest Pathophysiol 5:359-65.
104. Gootenberg DB, Paer JM, Luevano JM, Kwon DS. 2017. HIV-associated changes in the enteric microbial community: potential role in loss of homeostasis and development of systemic inflammation. Curr Opin Infect Dis 30:31-43.
105. Troseid M, Manner IW, Pedersen KK, Haissman JM, Kvale D, Nielsen SD. 2014. Microbial translocation and cardiometabolic risk factors in HIV infection. AIDS Res Hum Retroviruses 30:514-22.
106. Nobles CL, Maresso AW. 2011. The theft of host heme by Gram-positive pathogenic bacteria. Metallomics 3:788-96.
107. Anzaldi LL, Skaar EP. 2010. Overcoming the heme paradox: heme toxicity and tolerance in bacterial pathogens. Infect Immun 78:4977-89.
108. Gasche C, Lomer MC, Cavill I, Weiss G. 2004. Iron, anaemia, and inflammatory bowel diseases. Gut 53:1190-7.
109. Kaitha S, Bashir M, Ali T. 2015. Iron deficiency anemia in inflammatory bowel disease. World J Gastrointest Pathophysiol 6:62-72.
110. Williams HR, Cox IJ, Walker DG, North BV, Patel VM, Marshall SE, Jewell DP, Ghosh S, Thomas HJ, Teare JP, Jakobovits S, Zeki S, Welsh KI, Taylor-Robinson SD, Orchard TR. 2009. Characterization of inflammatory bowel disease with urinary metabolic profiling. Am J Gastroenterol 104:1435-44.
111. Williams HR, Cox IJ, Walker DG, Cobbold JF, Taylor-Robinson SD, Marshall SE, Orchard TR. 2010. Differences in gut microbial metabolism are responsible for reduced hippurate synthesis in Crohn's disease. BMC Gastroenterol 10:108.
112. Radogna F, Diederich M, Ghibelli L. 2010. Melatonin: a pleiotropic molecule regulating inflammation. Biochem Pharmacol 80:1844-52.

113. Paulose JK, Wright JM, Patel AG, Cassone VM. 2016. Human Gut Bacteria Are Sensitive to Melatonin and Express Endogenous Circadian Rhythmicity. PLoS One 11:e0146643.
114. Paiva CN, Bozza MT. 2014. Are reactive oxygen species always detrimental to pathogens? Antioxid Redox Signal 20:1000-37.
115. Morgan XC, Tickle TL, Sokol H, Gevers D, Devaney KL, Ward DV, Reyes JA, Shah SA, LeLeiko N, Snapper SB, Bousvaros A, Korzenik J, Sands BE, Xavier RJ, Huttenhower C. 2012. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. Genome biology 13:R79.
116. Preston A, Mandrell RE, Gibson BW, Apicella MA. 1996. The lipooligosaccharides of pathogenic gram-negative bacteria. Crit Rev Microbiol 22:139-80.
117. Cardenas L, Dominguez J, Santana O, Quinto C. 1996. The role of the nodI and nodJ genes in the transport of Nod metabolites in Rhizobium etli. Gene 173:183-7.
118. Vazquez M, Santana O, Quinto C. 1993. The NodL and NodJ proteins from Rhizobium and Bradyrhizobium strains are similar to capsular polysaccharide secretion proteins from gram-negative bacteria. Mol Microbiol 8:369-77.
119. Funderburg NT, Stubblefield Park SR, Sung HC, Hardy G, Clagett B, Ignatz-Hoover J, Harding CV, Fu P, Katz JA, Lederman MM, Levine AD. 2013. Circulating CD4(+) and CD8(+) T cells are activated in inflammatory bowel disease and are associated with plasma markers of inflammation. Immunology 140:87-97.
120. Integrative HMPRNC. 2014. The Integrative Human Microbiome Project: dynamic analysis of microbiome-host omics profiles during periods of human health and disease. Cell Host Microbe 16:276-89.
121. Human Microbiome Project C. 2012. Structure, function and diversity of the healthy human microbiome. Nature 486:207-14.
122. Sharpton TJ. 2017. A high-throughput and modular workflow to quality control shotgun metagenomic DNA sequence libraries, https://zenodo.org/record/834783 - .WYNwP9MrLAw.
123. Schmieder R, Edwards R. 2011. Quality control and preprocessing of metagenomic datasets. Bioinformatics 27:863-4.
124. Rotmistrovsky K, Agarwala R. 2011. BMTagger: Best Match Tagger for removing human reads from metagenomics datasets. doi:citeulike-article-id:9207261.
125. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:119.
126. Zhao Y, Tang H, Ye Y. 2012. RAPSearch2: a fast and memory-efficient protein similarity search tool for next-generation sequencing data. Bioinformatics 28:125-6.
127. Nayfach S, Pollard KS. 2015. Average genome size estimation improves comparative metagenomics and sheds light on the functional ecology of the human microbiome. Genome Biol 16:51.
128. Roberts DW. 2016. labdsv: Ordination and Multivariate Analysis for Ecology, https://cran.r-project.org/web/packages/labdsv/.
129. Jari Oksanen FGB, Michael Friendly, Roeland Kindt, Pierre Legendre, Dan McGlinn, Peter R. Minchin, R. B. O'Hara, Gavin L. Simpson, Peter Solymos, M. Henry H. Stevens, Eduard Szoecs, Helene Wagner. 2017. vegan: Community Ecology Package, https://cran.r-project.org/web/packages/vegan/.
130. Chen J, Bittinger K, Charlson ES, Hoffmann C, Lewis J, Wu GD, Collman RG, Bushman FD, Li H. 2012. Associating microbiome composition with environmental covariates using generalized UniFrac distances. Bioinformatics 28:2106-13.

131. Yarza P, Ludwig W, Euzeby J, Amann R, Schleifer KH, Glockner FO, Rossello-Mora R. 2010. Update of the All-Species Living Tree Project based on 16S and 23S rRNA sequence analyses. Syst Appl Microbiol 33:291-9.
132. Zhang Y. 2013. Likelihood-based and Bayesian methods for Tweedie compound Poisson linear mixed models. Statistics and Computing 23:743-757.
133. Ye Y, Doak TG. 2009. A parsimony approach to biological pathway reconstruction/inference for genomes and metagenomes. PLoS Comput Biol 5:e1000465.
134. Hejblum BP, Skinner J, Thiebaut R. 2015. Time-Course Gene Set Analysis for Longitudinal Gene Expression Data. PLoS Comput Biol 11:e1004310.
135. Wu S, Wu H. 2013. More powerful significant testing for time course gene expression data using functional principal component analysis approaches. BMC Bioinformatics 14:6.
136. Lenth RV. 2016. Least-Squares Means: The R Package lsmeans. J Stat Softw 69:33.

Supplemental Files

Additional File 1: Figure S1. Longitudinal phenotypic monitoring of T cell activation in PBMC of WT and DNR "bleeder" mice. Representative FACS plots from WT bleeder #89 and DNR bleeder #85 at indicated time points. The gate is set on $CD44^{hi}$ population to show the percent of activated effector and memory T cells over time.

Additional File 2: Figure S2. Phenotypic monitoring of CD8 T cell activation in WT and DNR "pooper" mice post disease onset. Representative FACS plots from spleen or mesenteric lymph node (MLN) isolated from WT pooper #90 and DNR pooper #83 at week 15. Fraction of activated ($CD44^{hi}$) and effector ($CD44^{hi}$ $KLRG\text{-}1^{hi}$) CD8+ T cells are shown.

Additional File 3: Table S1. Project metadata

Additional File 4: Figure S3. The functional alpha-diversity of the gut microbiome as measured by Shannon entropy differentially varies over time across cohorts (left). In fact, there is significantly greater variation in the Shannon entropy between lines before and after disease onset (right).

Additional File 5: Text S1. Modeling methods

Additional File 6: Table S2. KEGG modules that exhibit significant group by time interaction coefficients, indicating that they differentially diversify over time between the two lines.

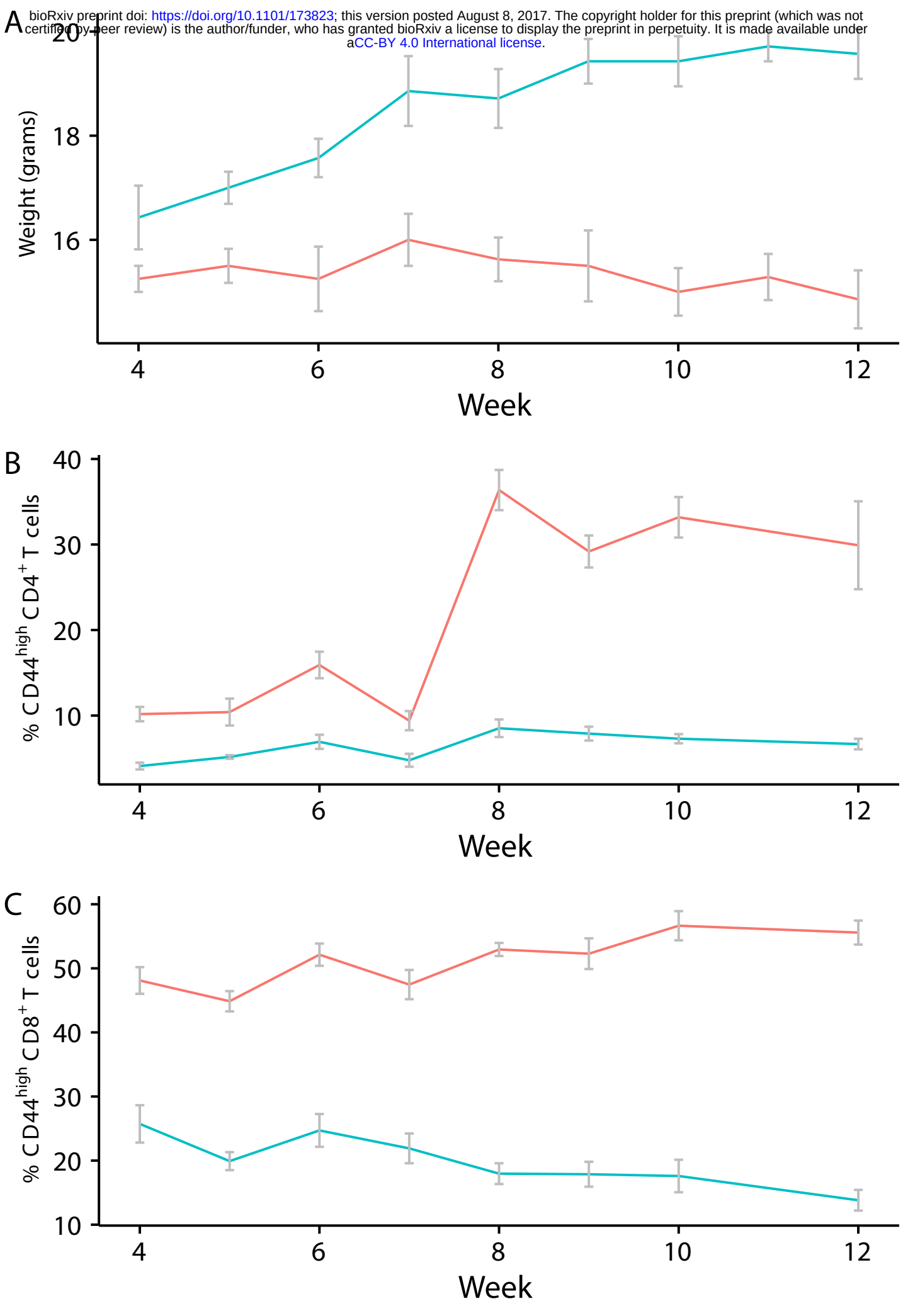
Additional File 7: Table S3 - KEGG Modules with significant interactions in the segmented GLMM analysis

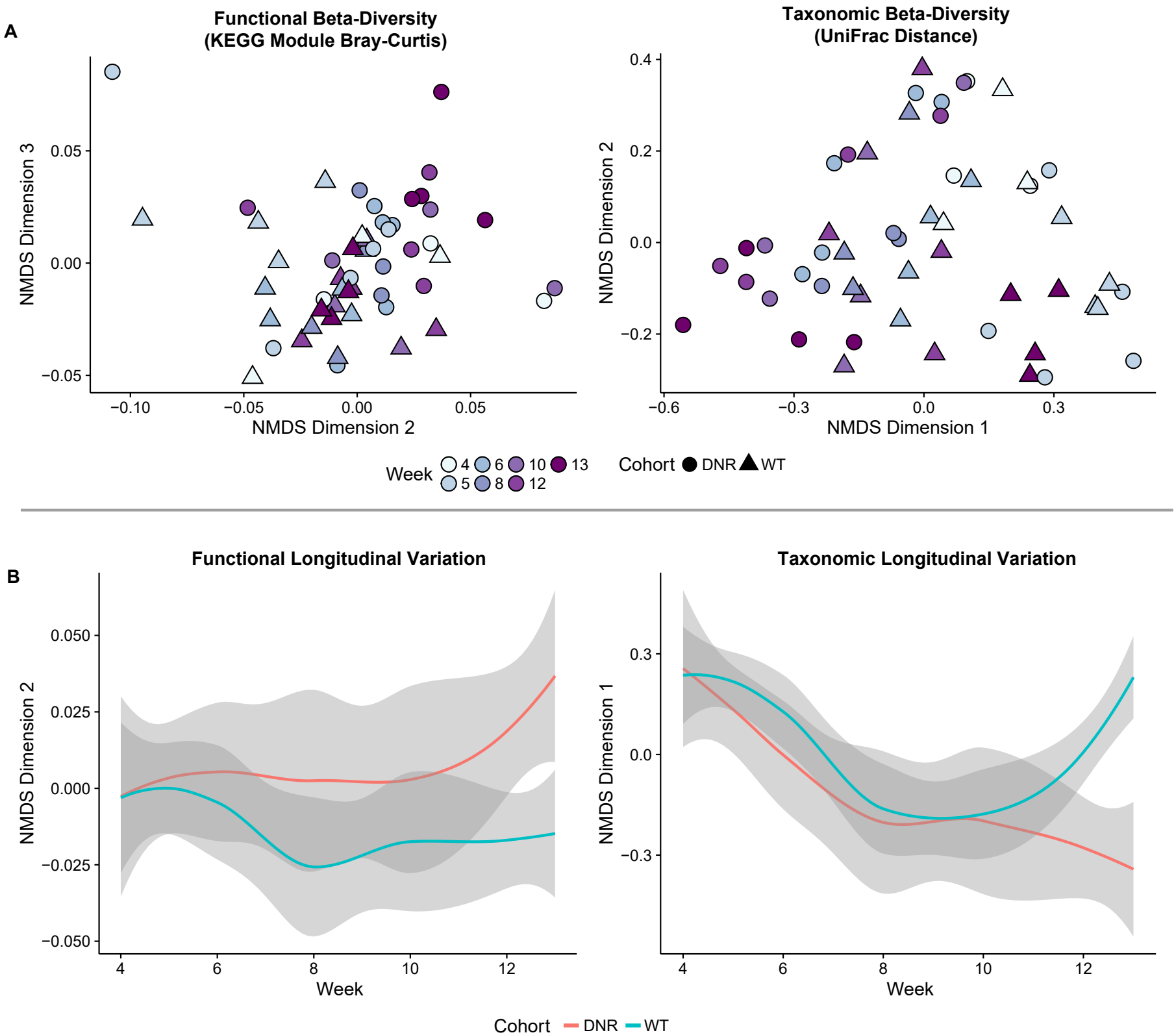
Additional File 8: Figure S4 - Two different analyses potentially explain the taxonomic origins of the lipooligosaccharide transport KOs that were observed. (a) Species identities of KEGG ortholog

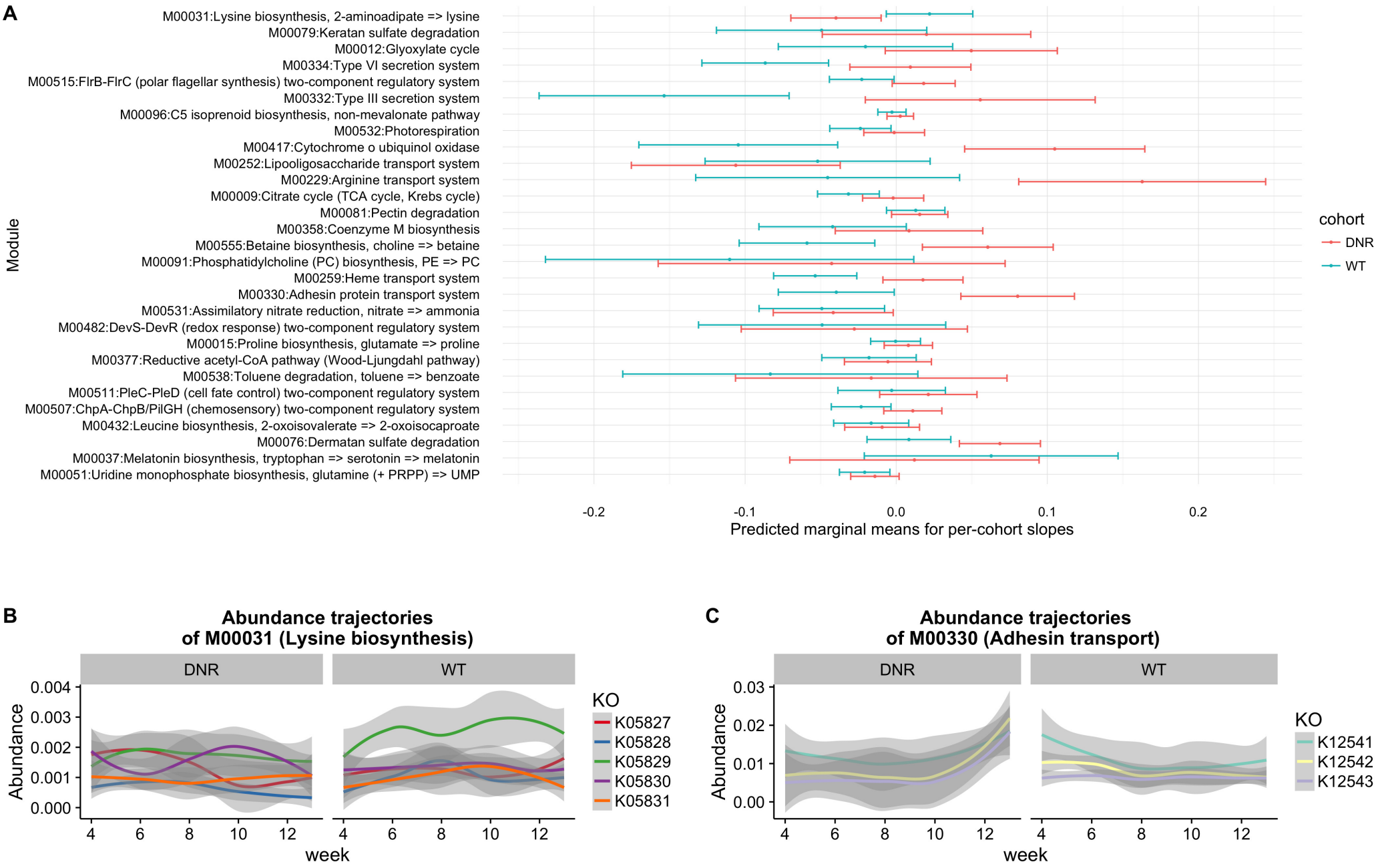
sequences that recruited reads when generating the relevant KO abundances. (b) Distance covariance values trajectories of K09694 and K09694 and all species, with asterisks marking those dCov values that were significantly non-zero after B-H multiple testing correction.

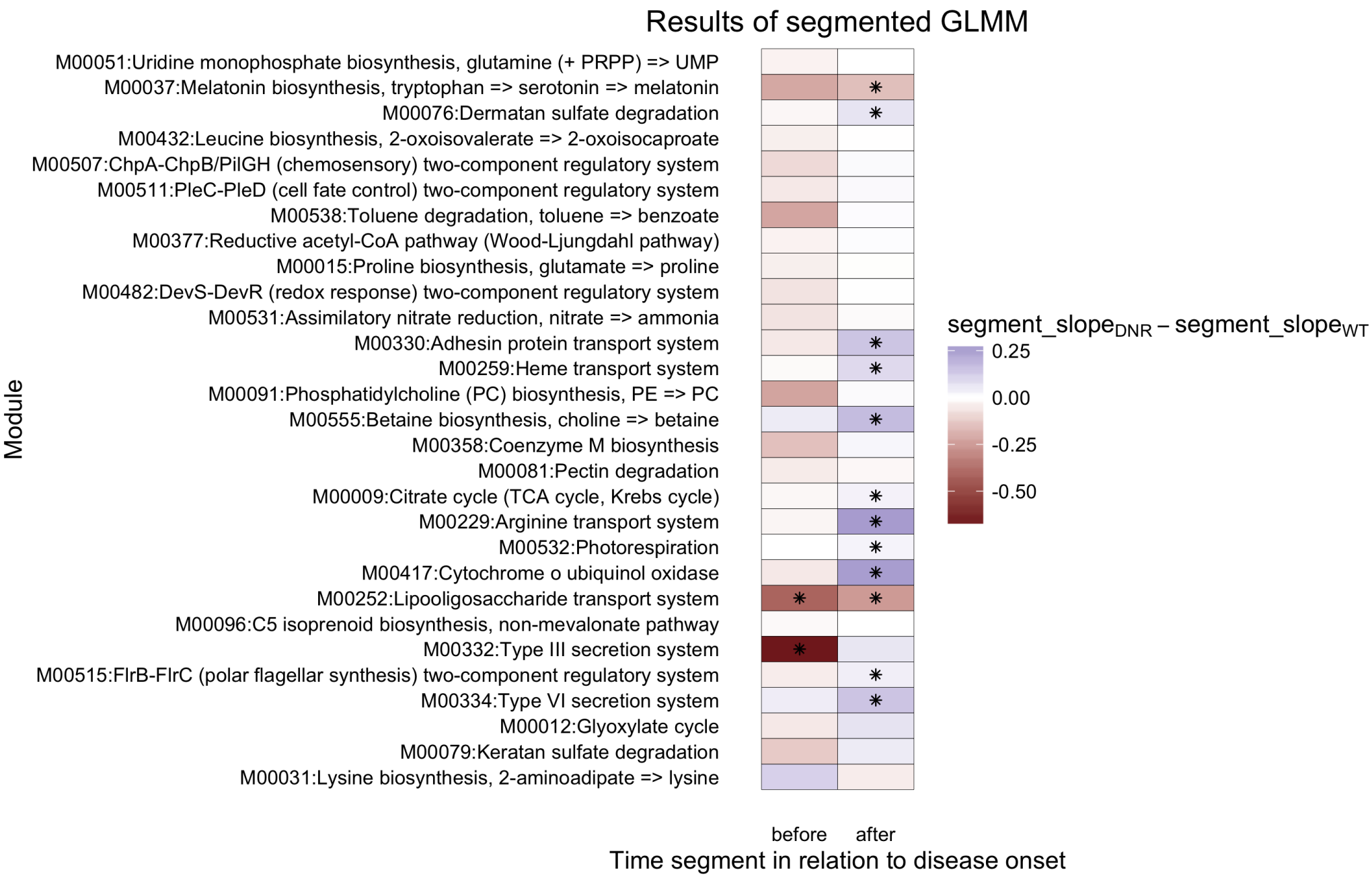
Additional File 9: Table S4 – KEGG modules with significant intercepts, indicating that they exhibited significantly different abundances between lines at the initial time point.

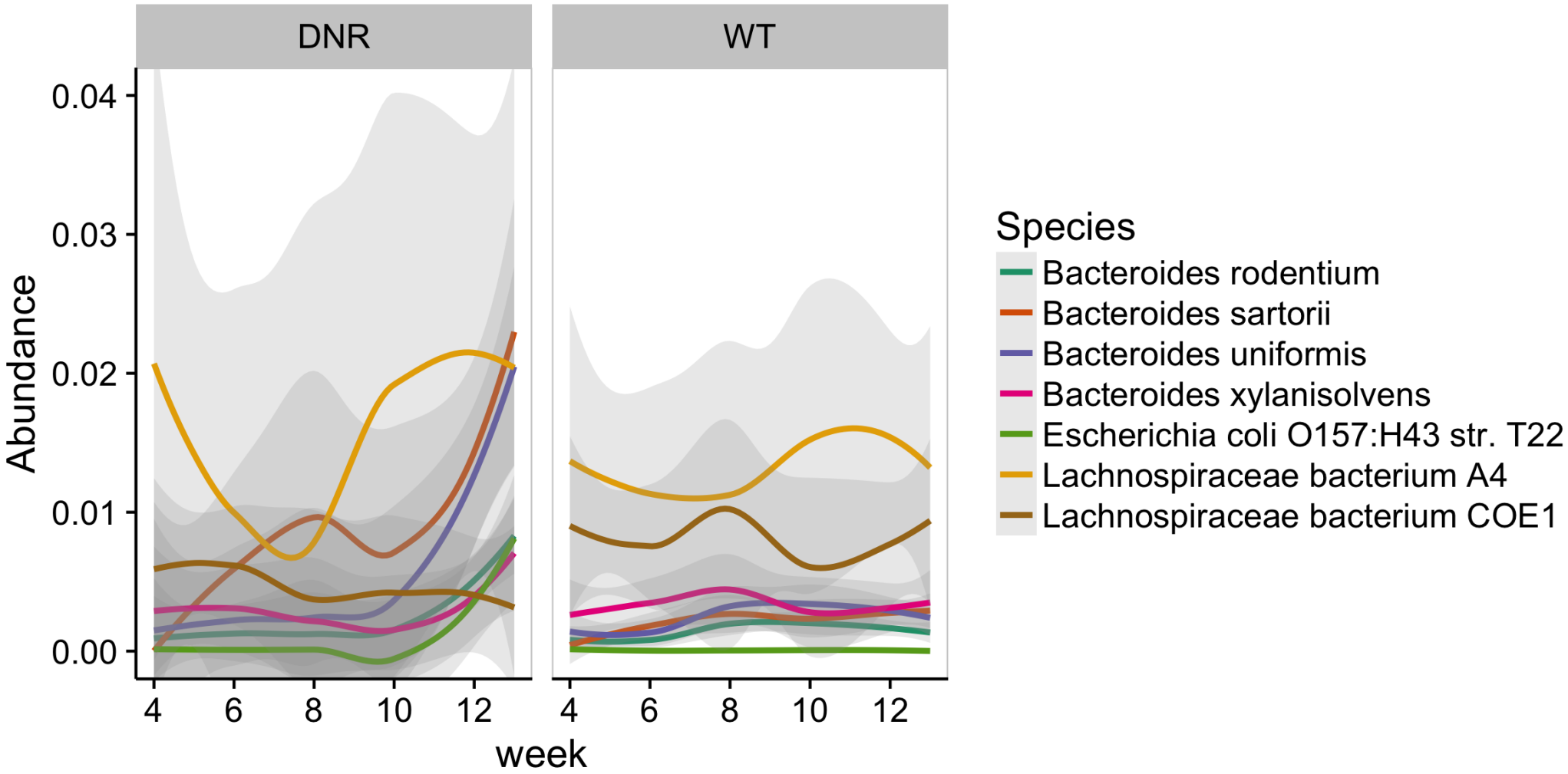
Additional File 10: Figure S5 - Distributions of F-statistics computed by the DISCO method for KO and species vectors within module and genus groupings respectively, compared to values from permuted groupings. Kolmogorov-Smirnov tests show significant difference between real and permuted distributions in the functional groupings, but no significant difference in the taxonomic groupings.




A
Functional Beta-Diversity
(KEGG Module Bray-Curtis)
NMDS Dimension 3
NMDS Dimension 2
Taxonomic Beta-Diversity
(UniFrac Distance)
NMDS Dimension 2
NMDS Dimension 1
Week 4 5 6 8 10 12 13
Cohort DNR WT
B
Functional Longitudinal Variation
NMDS Dimension 2
Week
Taxonomic Longitudinal Variation
NMDS Dimension 1
Week
Cohort DNR WT


A
M00031:Lysine biosynthesis, 2-aminoadipate => lysine
M00079:Keratan sulfate degradation
M00012:Glyoxylate cycle
M00334:Type VI secretion system
M00515:FlrB-FlrC (polar flagellar synthesis) two-component regulatory system
M00332:Type III secretion system
M00096:C5 isoprenoid biosynthesis, non-mevalonate pathway
M00532:Photorespiration
M00417:Cytochrome o ubiquinol oxidase
M00252:Lipooligosaccharide transport system
M00229:Arginine transport system
M00009:Citrate cycle (TCA cycle, Krebs cycle)
M00081:Pectin degradation
M00358:Coenzyme M biosynthesis
M00555:Betaine biosynthesis, choline => betaine
M00091:Phosphatidylcholine (PC) biosynthesis, PE => PC
M00259:Heme transport system
M00330:Adhesin protein transport system
M00531:Assimilatory nitrate reduction, nitrate => ammonia
M00482:DevS-DevR (redox response) two-component regulatory system
M00015:Proline biosynthesis, glutamate => proline
M00377:Reductive acetyl-CoA pathway (Wood-Ljungdahl pathway)
M00538:Toluene degradation, toluene => benzoate
M00511:PleC-PleD (cell fate control) two-component regulatory system
M00507:ChpA-ChpB/PilGH (chemosensory) two-component regulatory system
M00432:Leucine biosynthesis, 2-oxoisovalerate => 2-oxoisocaproate
M00076:Dermatan sulfate degradation
M00037:Melatonin biosynthesis, tryptophan => serotonin => melatonin
M00051:Uridine monophosphate biosynthesis, glutamine (+ PRPP) => UMP
Module
-0.2
-0.1
0.0
0.1
0.2
Predicted marginal means for per-cohort slopes
cohort
DNR
WT
B
Abundance trajectories
of M00031 (Lysine biosynthesis)
DNR
WT
0.004
0.003
0.002
0.001
0.000
Abundance
4
6
8
10
12
week
KO
K05827
K05828
K05829
K05830
K05831
C
Abundance trajectories
of M00330 (Adhesin transport)
DNR
WT
0.03
0.02
0.01
0.00
Abundance
4
6
8
10
12
week
KO
K12541
K12542
K12543


Results of segmented GLMM
M00051:Uridine monophosphate biosynthesis, glutamine (+ PRPP) => UMP
M00037:Melatonin biosynthesis, tryptophan => serotonin => melatonin
M00076:Dermatan sulfate degradation
M00432:Leucine biosynthesis, 2-oxoisovalerate => 2-oxoisocaproate
M00507:ChpA-ChpB/PilGH (chemosensory) two-component regulatory system
M00511:PleC-PleD (cell fate control) two-component regulatory system
M00538:Toluene degradation, toluene => benzoate
M00377:Reductive acetyl-CoA pathway (Wood-Ljungdahl pathway)
M00015:Proline biosynthesis, glutamate => proline
M00482:DevS-DevR (redox response) two-component regulatory system
M00531:Assimilatory nitrate reduction, nitrate => ammonia
M00330:Adhesin protein transport system
M00259:Heme transport system
M00091:Phosphatidylcholine (PC) biosynthesis, PE => PC
M00555:Betaine biosynthesis, choline => betaine
M00358:Coenzyme M biosynthesis
M00081:Pectin degradation
M00009:Citrate cycle (TCA cycle, Krebs cycle)
M00229:Arginine transport system
M00532:Photorespiration
M00417:Cytochrome o ubiquinol oxidase
M00252:Lipooligosaccharide transport system
M00096:C5 isoprenoid biosynthesis, non-mevalonate pathway
M00332:Type III secretion system
M00515:FlrB-FlrC (polar flagellar synthesis) two-component regulatory system
M00334:Type VI secretion system
M00012:Glyoxylate cycle
M00079:Keratan sulfate degradation
M00031:Lysine biosynthesis, 2-aminoadipate => lysine
Module
segment_slope$_{DNR}$ – segment_slope$_{WT}$
0.25
0.00
-0.25
-0.50
before
after
Time segment in relation to disease onset


DNR
WT
0.04
0.03
0.02
0.01
0.00
Abundance
4
6
8
10
12
week
Species
Bacteroides rodentium
Bacteroides sartorii
Bacteroides uniformis
Bacteroides xylanisolvens
Escherichia coli O157:H43 str. T22
Lachnospiraceae bacterium A4
Lachnospiraceae bacterium COE1

Species ID	p-value	q-value	Species Name	WT area under LOESS curve	DNR area under LOESS curve
54642	0	0	*Bacteroides sartorii*	0.01992	0.07699
57185	0	0	*Bacteroides xylanisolvens*	0.03051	0.02506
57318	0	0	*Bacteroides uniformis*	0.02297	0.04506
58110	0	0	*Escherichia coli O157:H43 str. T22*	5.35E-04	0.007523
59684	0.0001	0.0025	*Lachnospiraceae bacterium COE1*	0.07203	0.04213
59708	0	0	*Bacteroides rodentium*	0.0136	0.01986
61442	0.0013	0.02786	*Lachnospiraceae bacterium A4*	0.119	0.1348