Fecal microbiota signatures are associated with response to Ustekinumab therapy among Crohn's Disease patients

Running title: Microbiota of Ustekinumab-treated Crohn's subjects.

Matthew K. Doherty[1], Tao Ding[1α], Charlie Koumpouras[1], Shannon E. Telesco[2], Calixte Monast[2], Anuk Das[2], Carrie Brodmerkel[2], and Patrick D. Schloss[1†]

† To whom correspondence should be addressed: Patrick D. Schloss, pschloss@umich.edu

1. Department of Microbiology and Immunology, University of Michigan, Ann Arbor, MI, USA

2. Janssen Pharmaceutical Companies of Johnson & Johnson, Spring House, PA, USA

α Currently at Department of Biology, New York University, New York, NY, USA.

Abstract

The fecal microbiota is a rich source of biomarkers that have previously been shown to be predictive of numerous disease states. Less well studied is whether these biomarkers can be predictive of response to therapy. This study sought to predict the therapeutic response of Ustekinumab (UST) treated Crohn's disease (CD) patients. Using stool samples collected over the course of 22 weeks, the composition of these subjects' fecal bacterial communities was characterized by sequencing the 16S rRNA gene. Subjects in remission could be distinguished from those with active disease 6 weeks after treatment using Random Forest models trained on subjects' baseline microbiota and clinical data (AUC = 0.844, specificity = 0.831, sensitivity = 0.774). The most predictive OTUs that were ubiquitous among subjects were affiliated with *Faecalibacterium* and *Escherichia/Shigella*. Among subjects in remission 6 weeks after treatment, the median baseline community diversity was 1.7 times higher than treated subjects with active disease ($p = 0.020$). Their baseline community structures were also significantly different ($p = 0.017$). Two OTUs affiliated with *Faecalibacterium* ($p = 0.003$) and *Bacteroides* ($p = 0.022$) were significantly more abundant at baseline in subjects who were in remission 6 weeks after treatment than those with active CD. The diversity of UST treated clinical responders increased over the 22 weeks of the study, in contrast to nonresponsive subjects ($p = 0.01$). The observed baseline differences in fecal microbiota and changes due to therapeutic response support using the microbiota as a biomarker for predicting and monitoring a patient's response to UST. (word count= 244/250, TextWrangler)

Importance: CD is a global health concern, with increasing incidence and prevalence, causing large economic and health care impacts. Finding prognostic biomarkers that give clinicians the ability to predict response to CD treatment at diagnosis will reduce the time that subjects spend taking drugs that will not be beneficial. OTUs associated with remission after treatment induction, especially *Faecalibacterium*, could be biomarkers for successful UST treatment of TNF-α refractory CD patients. More broadly, these results suggest the fecal microbiota could be a use-

ful non-invasive biomarker for directing or monitoring the treatment of gastrointestinal diseases. (word count =94/150, TextWrangler)

Keywords: IBD, microbiome, biologics, prediction, biomarkers, remission, Stelara, machine learning

Introduction

The microbiome has been correlated with a variety of diseases and has shown promise as a predictive tool for disease outcome for gingivitis (1), cardiovascular disease (2), *Clostridium difficile* infection (3, 4), and colorectal cancer (5, 6). Additionally, the microbiome has been shown to alter the efficacy of vaginal microbicides (7), cardiac drugs (8), and cancer treatments (9, 10). These results demonstrate that it is possible to use biomarkers from within the microbiome to predict response to therapeutics. In relation to inflammatory bowel disease (IBD), previous studies have shown that the bacterial gut microbiota correlates with disease severity in new-onset, pediatric Crohn's disease (CD) patients (11, 12). Additionally, recent studies suggest the gut microbiota could be used to predict clinical response to treatment in individuals with IBD (13, 14). It remains to be determined, however, whether the composition of the fecal gut microbiota can predict and monitor response to CD therapy. Considering the involvement of the immune system and previous evidence for involvement of the microbiome, we hypothesize that response to immunological CD therapy can be predicted using microbiome data.

CD is a global health concern causing large economic and health care impacts (15, 16). The disease is characterized by patches of ulceration and inflammation along the entire gastrointestinal tract, with most cases involving the ileum and colon. Currently, individuals with CD are treated based on disease location and risk of complications using escalating immunosuppressive treatment, and/or surgery, with the goal of achieving and sustaining remission (17, 18). Faster induction of remission following diagnosis reduces the risk of irreversible intestinal damage and disability (18–20). Ideally, clinicians would be able to determine personalized treatment options for CD patients at diagnosis that would result in faster achievement of remission (21). Therefore, recent research has been focused on identifying noninvasive biomarkers to monitor CD severity and predict therapeutic response (22–24).

The precise etiology of CD remains unknown, but host genetics, environmental exposure, and the gut microbiome appear to be involved (15, 25). Individuals with CD have reduced microbial diver-

sity in their guts, compared to healthy individuals, with a lower relative abundance of *Firmicutes* and an increased relative abundance of *Enterobacteraciae* and *Bacteroides* (11, 26–29). Additionally, genome-wide association studies of individuals with CD identified several susceptibility loci including loci involved in the IL-23 signaling pathway, which could impact the gut microbiota composition and function (17, 26, 30–33). If the fecal microbiota can be used to monitor disease severity and predict response to specific treatment modalities, then clinicians could use it as a noninvasive tool for prescribing therapies that result in faster remission (34).

The FDA recently approved Ustekinumab (UST), a monoclonal antibody directed against the shared p40 subunit of IL-12 and IL-23, for the treatment of CD (18, 35–37). Given the potential impact of IL-23 on the microbiota (30–33), we hypothesized that response to UST could be predicted or influenced by differences in subjects' gut microbiota and that UST treatment may alter the fecal microbiota. The effects of biologic treatment of IBD on the microbiota are not yet well described, but are hypothesized to be indirect, as these drugs act on host factors (17). We analyzed the fecal microbiota of subjects who participated in a double-blinded, placebo-controlled Phase II clinical trial that demonstrated the safety and efficacy of UST for treating CD (35). The original study found that UST induction treatment had an increased rate of response as well as increased rates of response and remission with UST maintenance therapy, compared to placebo. We quantified the association between the fecal microbiota and disease severity. Finally, we tested whether clinical responders had a microbiota that was distinct from non-responders and whether the fecal microbiota changed in subjects treated with UST using 16S rRNA gene sequence data from these subjects' stool samples. Our study demonstrates that these associations are useful in predicting and monitoring UST treatment outcome and suggest the fecal microbiota may be a broadly useful source of biomarkers for predicting response to treatment.

Results

Study design

We characterized the fecal microbiota in a subset of anti-TNF-α refractory CD patients, with moderate to severe CD, who took part in a randomized, double-blinded, placebo-controlled phase 2b clinical trial that demonstrated the efficacy of UST in treating CD (35). Demographic and baseline disease characteristics of this subset are summarized in Table 1. Subjects were randomly assigned to a treatment group in the induction phase of the study and were re-randomized into maintenance therapy groups 8 weeks after induction based on their response (Figure 1A). In the current study, response was defined as a decrease in a subject's initial Crohn's Disease Activity Index (CDAI) greater than 30%. Remission was defined as a CDAI below 150 points. The CDAI is the standard instrument for evaluating clinical symptoms and disease activity in CD (38, 39). The CDAI weights patient reported stool frequency, abdominal pain, and general well being over a week, in combination with weight change, hematocrit, opiate usage for diarrhea, and the presence of abdominal masses or other complications to determine the disease severity score (38, 39). Subjects provided stool samples at baseline (screening) and at 4, 6, and 22 weeks after induction for analysis using 16S rRNA gene sequencing (Figure 1B).

Prediction of remission following treatment

We investigated whether the composition of the baseline fecal microbiota could predict therapeutic remission (CDAI $<$ 150) 6 weeks after induction. To test this hypothesis, we generated Random Forest (RF) models to predict which subjects would be in remission 6 weeks after induction based on the relative abundance of the fecal microbiota at baseline, clinical metadata at baseline, and the combination of microbiota and clinical data. We determined the optimal model based the largest area under the curve (AUC) of the receiver operating characteristic (ROC) curve for the RF model (6, 40). Clinical data included components of the CDAI, biomarkers for inflammation, and subject metadata described further in the methods section. We trained these models using 232 baseline stool samples from subjects induced with UST. Clinical data alone resulted in an AUC of 0.616 (specificity = 0.801, sensitivity = 0.452) (Figure 2A). Using only fecal microbiota data the model had an AUC of 0.838 (specificity = 0.766, sensitivity = 0.806). Finally, when combining

clinical metadata with the microbiota we achieved an AUC of 0.844 (specificity = 0.831, sensitivity = 0.774) for remission 6 weeks after induction. Prediction with clinical metadata alone did not perform as well as models using the baseline fecal microbiome ($p = 0.001$) or the combined model ($p = 0.001$); however, there was not a significant difference between the baseline fecal microbiota model and the combined model ($p = 0.84$).

Optimal predictors were determined based on their mean decrease in accuracy (MDA) in the ability of the model to classify remission from active CD (Figure 2B). The majority of OTUs identified as optimal predictors in our model for remission had low abundance. However, two OTUs were differentially abundant for subjects in remission 6 weeks after induction. The relative abundance of *Escherichia/Shigella* (OTU1) was lower in subjects in remission 6 weeks after induction (median = 1.07%, IQR = 0.033-3.7) compared to subjects with active CD (median = 4.13%, IQR = 0.667-15.4). Also, the relative abundance of *Faecalibacterium* (OTU7) was not only higher in subjects in remission 6 weeks after induction (median = 7.43%, IQR = 1.43-11.9) than subjects with active CD (median = 0.167%, IQR = 0-5.1), but it was also present prior to the start of treatment in every subject who was in remission 6 weeks after induction.

Prediction of response following treatment

We also hypothesized that the composition of the baseline fecal microbiota could predict therapeutic response (CDAI decrease >30%) 6 weeks after induction. To test this hypothesis, we again used RF models to classify responders from non-responders 6 weeks after induction. Clinical data alone resulted in an AUC of 0.651 (specificity = 0.545, sensitivity = 0.724) (Figure 2C). Using only microbiota data, the model predicted response with an AUC of 0.762 (specificity = 0.558, sensitivity = 0.882). When combining clinical metadata with the microbiome, the model predicted response with an AUC of 0.733 (specificity = 0.724, sensitivity = 0.684). These models were not significantly different in their ability to predict response ($p > 0.05$ for each comparison). Optimal predictors were again determined based on their MDA in the ability of the model to classify response (Figure 2D). Also, our baseline fecal microbiota model was significantly better

able to classify remission compared to response (p = 0.12), whereas this was not true for the combined model (p = 0.036).

Comparison of baseline microbiota based on clinical outcome

As our RF models identified OTUs abundant across our cohort that were important in classification of outcome, we further investigated differences in the baseline microbiota that could serve as potential biomarkers for successful UST treatment. We compared the baseline microbiota of all 306 subjects who provided a baseline sample based on treatment group and treatment outcome 6 weeks after induction. There was no significant difference in diversity based on response 6 weeks after induction, however the baseline β-diversity was significantly different by response (p = 0.018). No phyla were significantly different by treatment and response (Fig. S1) and no OTUs were significantly different based on UST response or among subjects receiving placebo for induction, regardless of response and remission status.

Subjects in remission 6 weeks after induction with UST had significantly higher baseline α-diversity based on the inverse Simpson diversity index than subjects with active CD (respective median values = 11.6 (IQR = 4.84-13.4), 6.95 (IQR = 4.25-11.8), p = 0.020). The baseline community structure was also significantly different based on remission status in subjects 6 weeks after induction (p = 0.017). Finally, 2 OTUs were significantly more abundant in subjects in remission 6 weeks after induction compared to subjects with active CD: *Bacteroides* (OTU19) (p = 0.022) and *Faecalibacterium* (OTU7) (p = 0.003) (Figure 3).

Variation in the baseline microbiota is associated with variation in clinical data

Based on the associations we identified between baseline microbial diversity and response, we hypothesized that there were associations between the microbiota and clinical variables at baseline that could support the use of the microbiota as a non-invasive biomarker for disease activity (34). To test this hypothesis, we compared the baseline microbiota with clinical data at baseline for all 306 samples provided at baseline (Supplemental Table 1). We observed small, but significant correlations for lower α-diversity correlating with higher CDAI (ρ = -0.161, p = 0.014), higher

frequency of loose stools per week (ρ = -0.193, p = 0.003), and longer disease duration (ρ = -0.225, p = 0.001). Corticosteroid use was associated with 1.45 times higher α-diversity (p = 0.001). No significant associations were observed between α-diversity and CRP, fecal calprotectin, or fecal lactoferrin. However, the β-diversity was significantly different based on CRP (p = 0.033), fecal calprotectin (p = 0.006), and fecal lactoferrin (p = 0.004). The β-diversity was also significantly different based on weekly loose stool frequency (p= 0.024), age (p = 0.033), the tissue affected (p = 0.004), corticosteroid use (p =0.010), and disease duration (p = 0.004). No significant differences in α or β diversity were observed for BMI, weight, or sex.

The diversity of the microbiota changes in UST responders

We tested whether treatment with UST altered the microbiota by performing a Friedman test comparing α-diversity, based on the inverse Simpson diversity index, at each time point within each treatment group based on the subject's response 22 weeks after induction. We included 48 subjects induced and maintained with UST (18 responders, 30 non-responders) and 14 subjects induced and maintained with placebo (8 responders, 6 non-responders), who provided samples at every time point (Figure 1). We saw no significant difference in the α-diversity over time in subjects who did not respond 22 weeks after induction, regardless of treatment, and in subjects who received placebo (Figure 4). However, the median α-diversity of responders 22 weeks after UST induction significantly changed over time (p = 0.01) having increased from baseline (median = 6.65, IQR = 4.6 - 9.24) to 4 weeks after UST induction(median = 9.33, IQR = 6.54 - 16.7), decreased from 4 to 6 weeks after induction (median = 8.42, IQR = 4.93 - 17.5), and was significantly higher than baseline (p $<$ 0.05) at 22 weeks after induction (median = 10.7, IQR = 5.49 - 14.6).

The microbiota after induction can distinguish between treatment outcomes

Having demonstrated the microbiome changes in subjects who responded to UST treatment, we hypothesized that the microbiota could be used to monitor response to UST therapy by classifying subjects based on disease activity (34). We again constructed RF classification models

to distinguish between subjects by UST treatment outcome based on their fecal microbiota 6 weeks after induction (6, 40). The study design resulted in only 75 week twenty-two stool samples from subjects induced and maintained with UST, so we focused our analysis on the 220 week 6 stool samples from subjects induced with UST. We were again better able to distinguish subjects in remission from subjects with active CD compared to responders from non-responders ($p = 0.005$; Figure 5A). Our model could classify response 6 weeks after induction using week 6 stool samples from subjects treated with UST with an AUC of 0.72 (sensitivity $= 0.563$, specificity $= 0.812$). For classifying subjects in remission from subjects with active CD 6 weeks after UST induction using week 6 stool samples, the model had an AUC of 0.866 (sensitivity $= 0.833$, specificity $= 0.832$). OTUs that were important for these classifications again included *Faecalibacterium* (OTU7), as well as *Blautia* (OTU124), *Clostridium XIVa* (OTU73), *Ruminococcaceae* (OTU53), and *Roseburia* (OTU12). These all had higher median relative abundance in subjects in remission 6 weeks after induction than those with active disease (Figure 5B).

Discussion

We sought to determine whether fecal microbiota can be used to identify patients who will respond to UST therapy and to gain a more detailed understanding of how UST treatment may affect the microbiota. We demonstrated that the microbiota could predict remission following UST therapy, compared to clinical metadata alone, in this unique cohort. We also found the fecal microbiota to be associated with CD severity metrics and treatment outcomes. Finally, we found that the microbiota of treated responders changed over time. These results helped further our understanding of the interaction between the human gut microbiota and CD in adult subjects with moderate-to-severe CD refractory to anti-TNF-α therapies.

The development of predictive models for disease or treatment outcome is anticipated to have a significant impact on clinical decision-making in health care (41). These models will help clinicians decide on the correct course of disease treatment or interventions for disease prevention

with their patients. Additionally, patients will benefit with more individualized care that will potentially reduce adverse effects and result in faster recovery, reduce expenses from ineffective therapies, or increase quality of life by preventing disease in patients with high risk.

Our predictive model revealed potential microbial biomarkers for successful UST therapy and allowed us to generate hypotheses about the biology of CD as it relates to the microbiome and UST response. *Faecalibacterium* frequently occurred in our models. It is associated with health, comprising up to 5% of the relative abundance in healthy individuals, and is generally rare in CD patients (26, 28, 42, 43). Each subject in remission 6 weeks after UST induction had measurable *Faecalibacterium* present at baseline. This supports the hypothesis that *Faecalibacterium* impacts CD pathogenesis. It may even be beneficial to administer *Faecalibacterium* as a probiotic during therapy. *Escherichia/Shigella* also occurred frequently in our models. This OTU is associated with inflammation and has been shown to be associated with CD (43). Many other taxa observed in our analysis had low abundance or were absent in the majority of subjects. However, in many cases these taxa are related and may serve similar ecologic and metabolic roles in the gut environment. We hypothesize that these microbes may have genes that perform similar metabolic functions. Performing metagenomics on stool samples in future studies, especially in patients who achieve remission, could reveal these functions, which could be further developed into a clinically useful predictive tool.

We were better able to predict whether a subject would go into remission rather than respond to treatment, as determined by CDAI score. We hypothesize that this was due to the relative nature of the response criteria compared to the threshold used to determine remission status. While the field appears to be moving away from CDAI and towards more objectively quantifiable patient reported outcomes and endoscopic verification of mucosal healing [(19); PRO ref], research is ongoing to discover less invasive and objectively quantifiable biomarkers [biomarkers refs].

We identified several associations between the microbiota and clinical variables that could impact how CD is monitored and treated in the future. Serum CRP, fecal calprotectin, and fecal lactofer-

rin are widely used as biomarkers to measure inflammation and CD severity. We found that the microbial community structure was different among subjects based on these markers. These results support the hypothesis that the fecal microbiota could function as a biomarker for measuring disease activity in patients, especially in concert with established inflammatory biomarkers (34, 44, 45). We also found that higher CDAI scores were associated with lower microbial diversity. This is consistent with other studies on the microbiota in individuals with CD compared to healthy individuals and studies looking at active disease compared to remission (11, 34, 46). However, the CDAI sub score of weekly stool frequency may have driven these differences (Supplementary Table 1), which is consistent with previous studies (47). We also observed differences in the microbial community structure based on disease localization, which is consistent with a study by Naftali et al (42). Our study also found that corticosteroid use impacts the composition of the human fecal microbiota, which is consistent with observations in mouse models (48). We also observed that longer disease duration is associated with a reduction in fecal microbial diversity. We hypothesize that prolonged disease duration and the associated inflammation results in the observed decrease in diversity.

Further research into fecal microbiota as a source of biomarkers for predicting therapeutic response could eventually allow for the screening of patients with stool samples at diagnosis to better inform treatment decisions for a wide range of diseases. For CD specifically, using the microbiota to predict response to specific treatment modalities could result in more personalized treatment and faster achievement of remission, thereby increasing patients' quality of life and reducing economic and health care impacts for CD patients. Our results showing that the α-diversity of clinical UST responders increased over time, in contrast to non-responsive patients, and our ability to classify patients in remission from those with active disease following UST treatment are again consistent with other studies suggesting the microbiota could be a useful biomarker for predicting or monitoring response to treatment (34). These predictive biomarkers will need to be validated using independent cohorts in future studies. Additionally, the positive and negative associations between the microbiota and CD allow us to predict the types of mechanisms most likely to alter the

microbiota in order to increase the likelihood of achieveing a therapeutic response or to monitor disease severity. Prior to the initiation of therapy, patients could have their fecal microbiome analyzed. Then the microbial community data could be used to direct the modification of a patient's microbiota prior to or during treatment with the goal of improving treatment outcomes. Since it has been shown that the microbiome can alter the efficacy of treatments for a variety of diseases (7–10), if fecal microbiota can be validated as biomarkers to non-invasively predict response to therapy, then patients and clinicians will be able to more rapidly ascertain effective therapies that result increased patient quality of life.

Methods

Study Design and Sample Collection

Previously, a randomized, double-blinded, placebo-controlled phase II clinical study of approximately 500 subjects assessed the safety and efficacy of UST for treating anti-TNF-α refractory, moderate to severe CD subjects (35) (Figure 1). Institutional review board approval was acquired at each participating study center and subjects provided written informed consent (35). Subject data was de-identified for our study. Participants provided a stool sample prior to the initiation of the study and were then divided into treatment groups. Additional stool samples were provided 4 weeks after induction. At 6 weeks after induction an additional stool sample was collected, subjects were scored for their response to UST based on CDAI, and then divided into groups receiving either subcutaneous injection of UST or placebo at weeks 8 and 16 as maintenance therapy. Clinical response was defined as a reduction from baseline CDAI score of 100 or more points or as remission in subjects with a baseline CDAI score between 220 to 248 points (35). Remission was defined as a CDAI below the threshold of 150. Finally, at 22 weeks subjects provided an additional stool sample and were then scored using CDAI for their response to therapy. Frozen fecal samples were shipped to the University of Michigan and stored at -80°C prior to DNA extraction.

DNA extraction and 16S rRNA gene sequencing

Microbial genomic DNA was extracted using the PowerSoil-htp 96 Well Soil DNA Isolation Kit (MoBio Laboratories) and an EPMotion 5075 pipetting system (5, 6). The V4 region of the 16S rRNA gene from each sample was amplified and sequenced using the Illumina MiSeq™ platform (45). Sequences were curated as described previously using the mothur software package (v.1.34.4) (49, 50). Briefly, we curated the sequences to reduce sequencing and PCR errors (51), aligned the resulting sequences to the SILVA 16S rRNA sequence database (52), and used UCHIME to remove any chimeric sequences (53). Sequences were clustered into operational

taxonomic units (OTU) at a 97% similarity cutoff using the average neighbor algorithm (54). All sequences were classified using a naive Bayesian classifier trained against the RDP training set (version 14) and OTUs were assigned a classification based on which taxonomy had the majority consensus of sequences within a given OTU (55).

Following sequence curation using the mothur software package (49), we obtained a median of 13,732 sequences per sample (IQR = 7,863-21,978). Parallel sequencing of a mock community had an error rate of 0.017%. To limit effects of uneven sampling, we rarefied the dataset to 3,000 sequences per sample. Samples from subjects that completed the clinical trial and had complete clinical metadata were included in our analysis. Of these samples, 306 were provided prior to treatment as well as 258 provided at week 4, 289 at week 6, and 205 at week 22 after treatment, for a total of 1,058 samples. All fastq files and the MIMARKS spreadsheet with de-identified clinical metadata are available at **SRA**.

Gut microbiota biomarker discovery and statistical analysis

R v.3.3.2 (2016-10-31) and mothur were used to analyze the data (56). To assess α-diversity, the inverse Simpson index was calculated for each sample in the dataset. Spearman correlation tests were performed to compare the inverse Simpson index and continuous clinical data. Wilcoxon rank sum tests were performed for pairwise comparisons and Kruskal-Wallis rank sum tests for comparisons with more than two groups (57, 58). To measure β-diversity, the distance between samples was calculated using the θYC metric, which takes into account the types of bacteria and their abundance to calculate the differences between the communities (59). These distance matrices were assessed for overlap between sets of communities using the non-parametric analysis of molecular variance (AMOVA) test as implemented in the adonis function from the vegan R package (v.2.4.3) (60). Changes in α-diversity over time based on week 22 response was assessed using a Friedman test on subjects who provided a sample at each time point (61). The Friedman test is a function in the stats R package (v.3.3.2). Multiple comparisons following a Friedman test were performed using the friedmanmc function in the pgirmess package (v.1.6.5) (62). Changes

in β-diversity over time by treatment group and response were assessed using the adonis function in vegan stratified by subject. We used the relative abundance of each OTU, α-diversity, age, sex, current medications, BMI, disease duration, disease location, fecal calprotectin, fecal lactoferrin, C-reactive protein, bowel stricture, and CDAI sub scores as input into our RF models constructed with the AUCRF R package (v.1.1) (63), to identify phylotypes or clinical variables that distinguish between various treatment and response groups, as well as to predict or determine response outcome (64). Optimal predictors were determined based on their mean decrease in accuracy (MDA) of the model to classify subjects. Differentially abundant OTUs and phyla were selected through comparison of clinical groups using Kruskal-Wallis and Wilcox tests, where appropriate, to identify OTUs/phyla where there was a p-value less than 0.05 following a Benjamini-Hochberg correction for multiple comparisons (65). Other R packages used in our analysis included ggplot2 v.2.2.1 (66), dplyr v.0.5.0 (67), pROC v.1.9.1 (68), knitr v.1.15.1 (69), gridExtra v.2.2.1 (70), devtools v.1.12.0 (71), knitcitations v.1.0.7 (72), scales v.0.4.1 (73), tidyr v.0.6.1 (74), Hmisc v.4.0.2 (75), and cowplot v.0.7.0 (76). A reproducible version of this analysis and manuscript are available at https://github.com/SchlossLab/Doherty_CDprediction_mBio_2017.

Funding information.

Janssen Research and Development provided financial and technical support for this study.

Tables

Table 1: Summary of clinical metadata of cohort at baseline

Clinical Variable	Treated	Placebo	Total
	n = 232	n = 74	n = 306
Age (years)	38 ± 13	40 ± 14	39 ± 13
Sex (% Male)	36.6	43.2	38.2
Race (% Caucasian)	91.8	93.2	92.2
Corticosteroid Use (%)	40.1	52.7	43.1
BMI (kg/m^2)	26 ± 6.7	25 ± 4.9	25 ± 6.3
Disease Duration (years)	12 ± 8.4	13 ± 10	12 ± 8.8
CDAI	330 ± 62	310 ± 69	320 ± 64
Bowel Stricture (%)	12.5	10.8	12.1
Tissue Involvement (%) Colon/Ileocolic/Ileal	28.9/51.3/19.8	24.3/39.2/36.5	27.8/48.4/23.9

Supplemental Table 1: Diversity differences based on clinical metadata of cohort at baseline

Clinical Variable	Correlation	Alpha-Diversity (p-value)	Beta-Diversity (p-value)
CDAI	ρ = -0.2	0.014	0.324
Loose Stool Frequency (per week)	ρ = -0.2	0.003	0.024
C-Reactive Protein (mg/L serum)	ρ = 0.06	0.394	0.033
Fecal Calprotectin (µg/g)	ρ = 0.08	0.254	0.006
Fecal Lactoferrin (µg/g)	ρ = 0.1	0.070	0.004
BMI	ρ = 0.07	0.299	0.277
Weight (kg)	ρ = 0.07	0.299	0.112
Age (years)	ρ = -0.05	0.472	0.033
Sex (F/M)	-	0.539	0.277
Corticosteroid Use (Y/N)	-	0.001	0.010
Disease Duration (years)	ρ = -0.2	0.001	0.004
Tissue Involvement	-	0.190	0.004

Figures

Figure 1: Experimental design as adapted from Sandborn et al 2012. (A) Participants were divided into treatment groups receiving placebo or UST by IV for induction. At week 8, subjects were divided into groups receiving either subcutaneous injection of UST or placebo at weeks 8 and 16 as maintenance therapy, based on response at week 6. Finally, at 22 weeks subjects were scored using CDAI for their response to therapy. (B) Stool sampling, treatment, and response evaluation time line. ↑, treatment administration; IV, intravenous; PE, primary endpoint; R, randomization; RR, re-randomization (only for subjects receiving UST induction therapy); SC, subcutaneous.

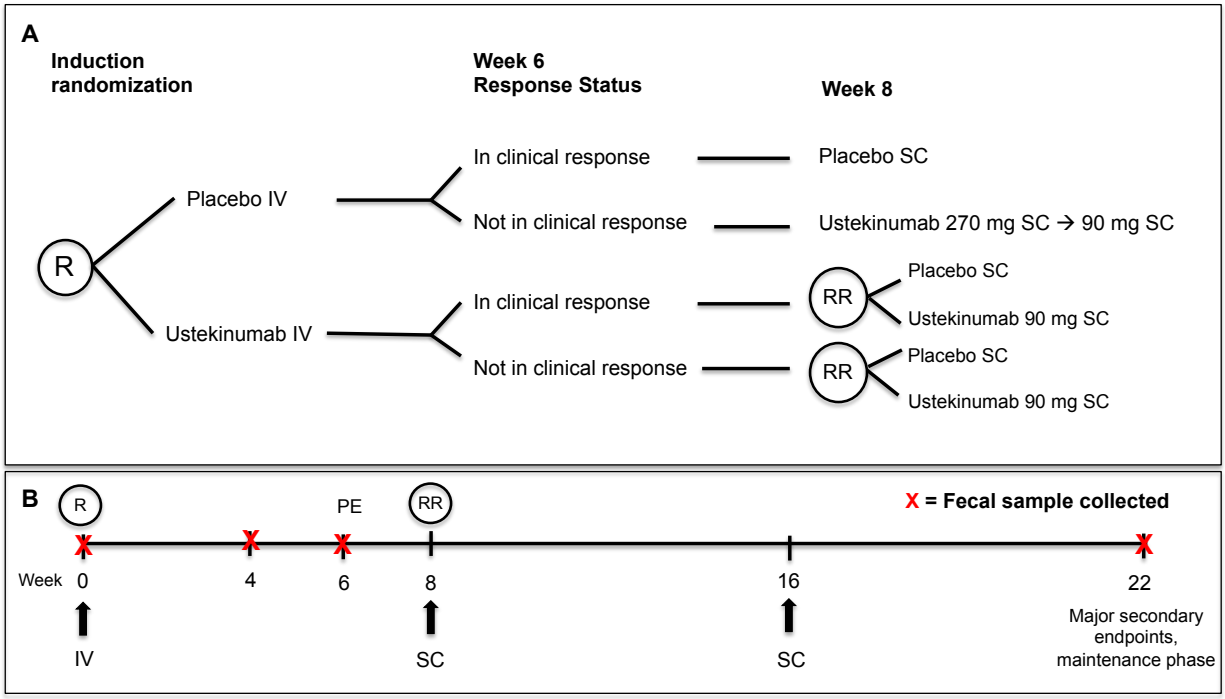
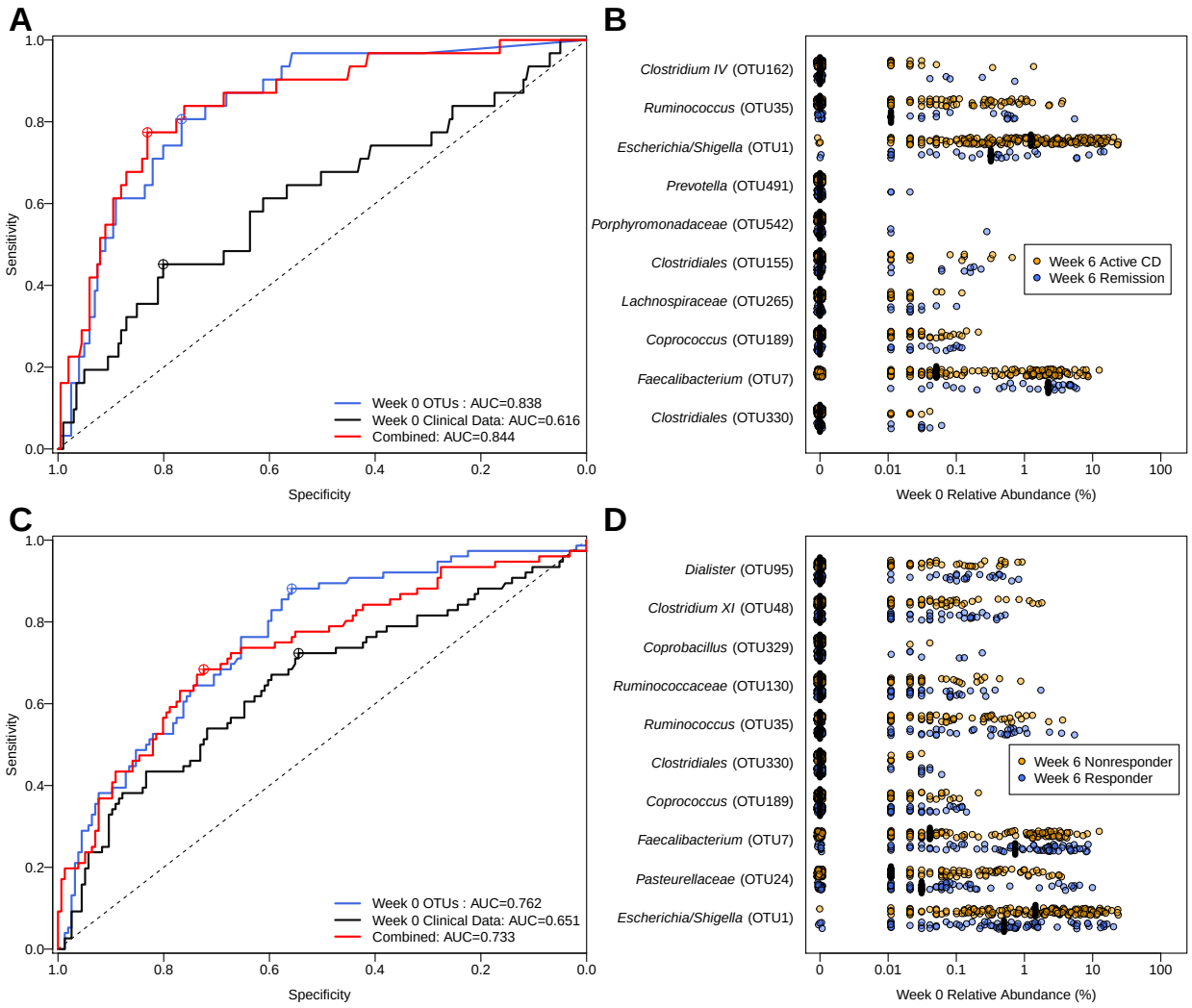


Figure 2: Prediction of week 6 treatment outcome in subjects treated with UST, using baseline samples Receiver operating characteristic (ROC) curves for (A) response and (C) remission using microbiota data (blue), clinical metadata (black), and a combined model (red). Top predictive OTUs for the microbiota model based on mean decrease in accuracy (MDA) for (B) response and (D) remission. Black bars represent the median relative abundance.



Supplemental Figure 1: Phyla from baseline stool samples in subjects treated with UST by week six outcome The relative abundance of each phylum in UST treated subjects were compared based on (A) response and (B) remission status using a Wilcoxon rank sum test and to identify phyla where there was a p-value less than 0.05 following a Benjamini-Hochberg correction for multiple comparisons. No comparisons were significant. Whiskers represent the range and boxes represent the 25-75% interquartile range of the median (black bar).

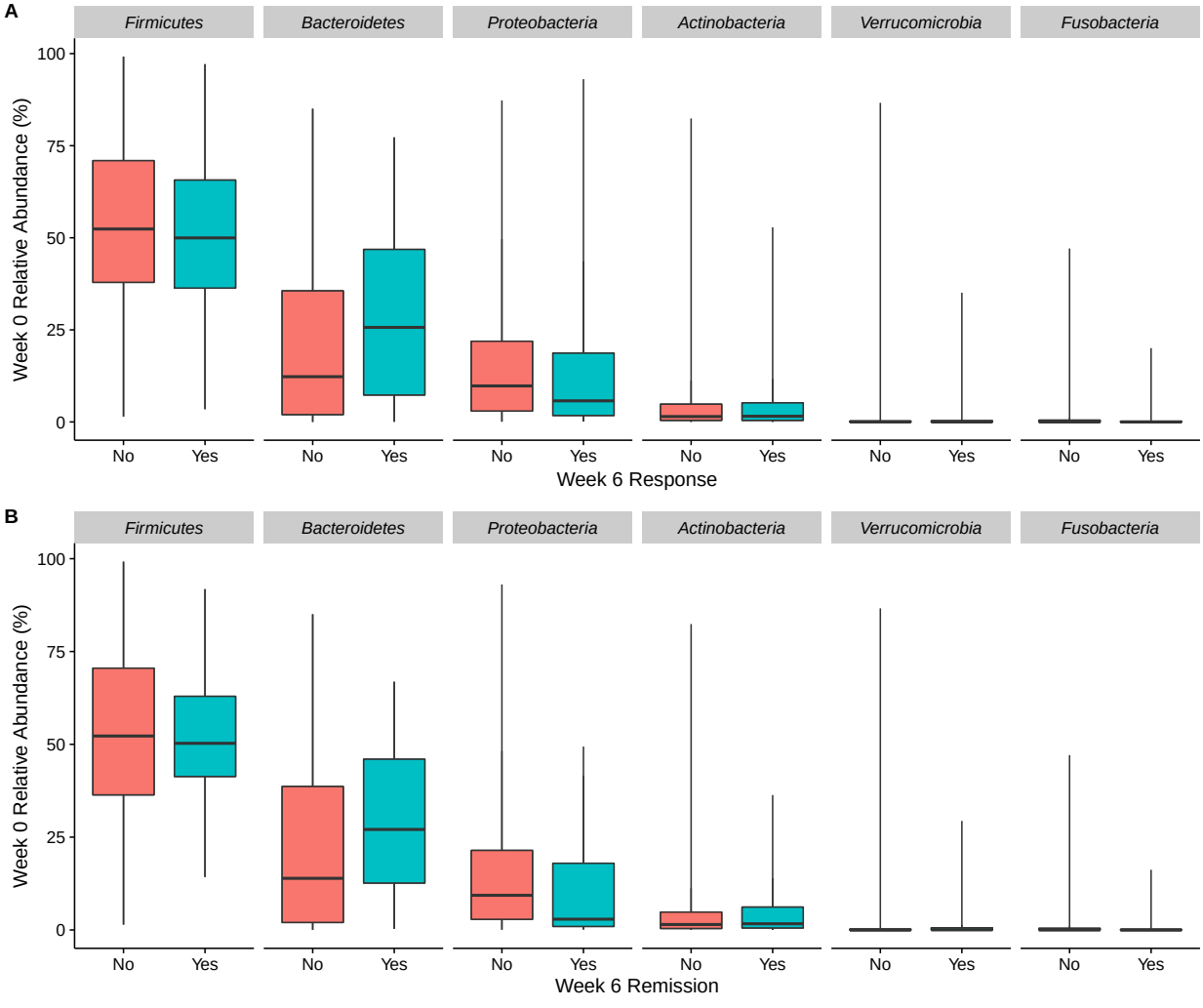


Figure 3: Differential taxa in baseline stool samples from subjects treated with UST, based on week six remission status The baseline relative abundance of each OTU was compared between subjects in remission and those with active CD 6 weeks after induction using a Wilcoxon rank sum test followed by a Benjamini-Hochberg correction for multiple comparisons. This identified 2 OTUs with significantly different relative abundance at baseline ($p < 0.05$). Black bars represent the median relative abundance.

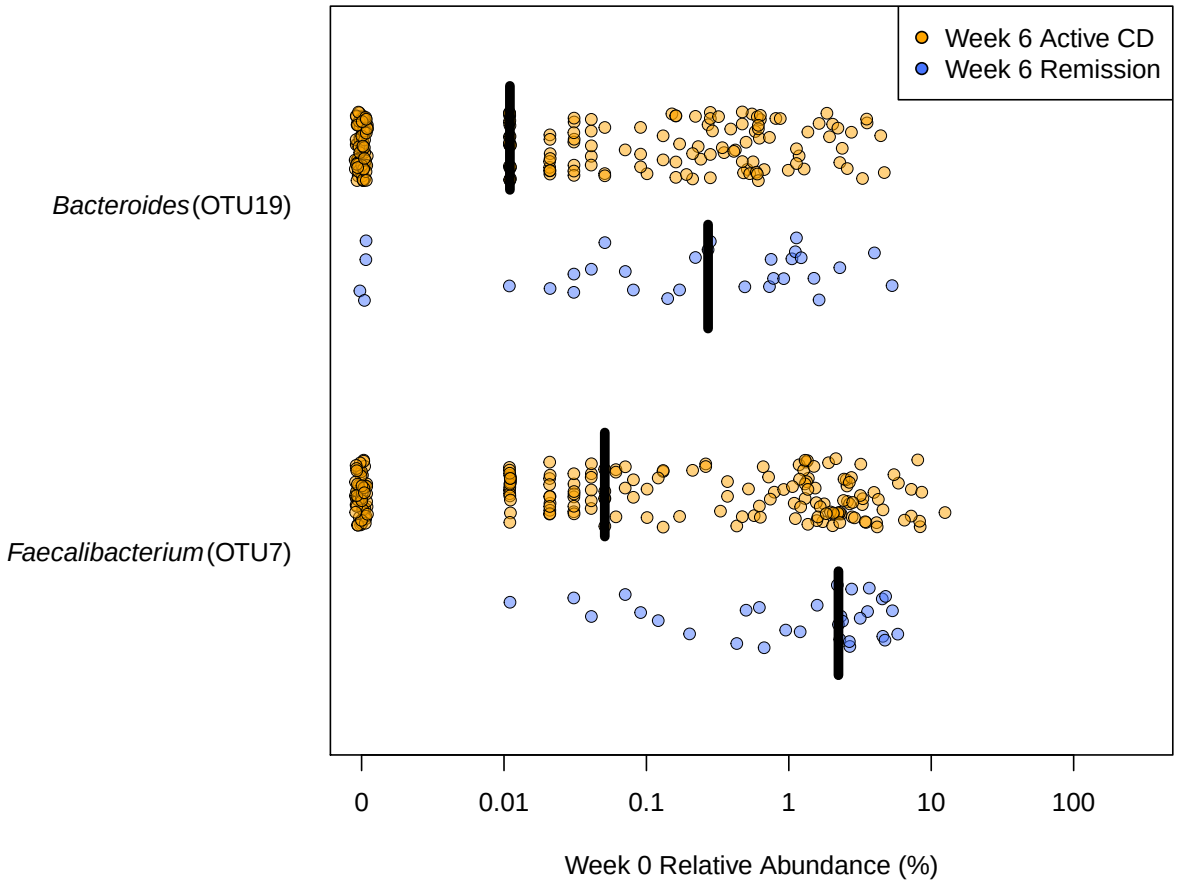


Figure 4: Change in alpha diversity over time by induction treatment and week 22 response status. The α-diversity of 48 subjects induced and maintained with UST and 14 subjects induced and maintained with placebo was assessed at each time point. Friedman test were performed within each treatment and responder group. Whiskers represent the range and boxes represent the 25-75% interquartile range of the median (black bar). * indicates week 22 is significantly different from baseline ($p < 0.05$).

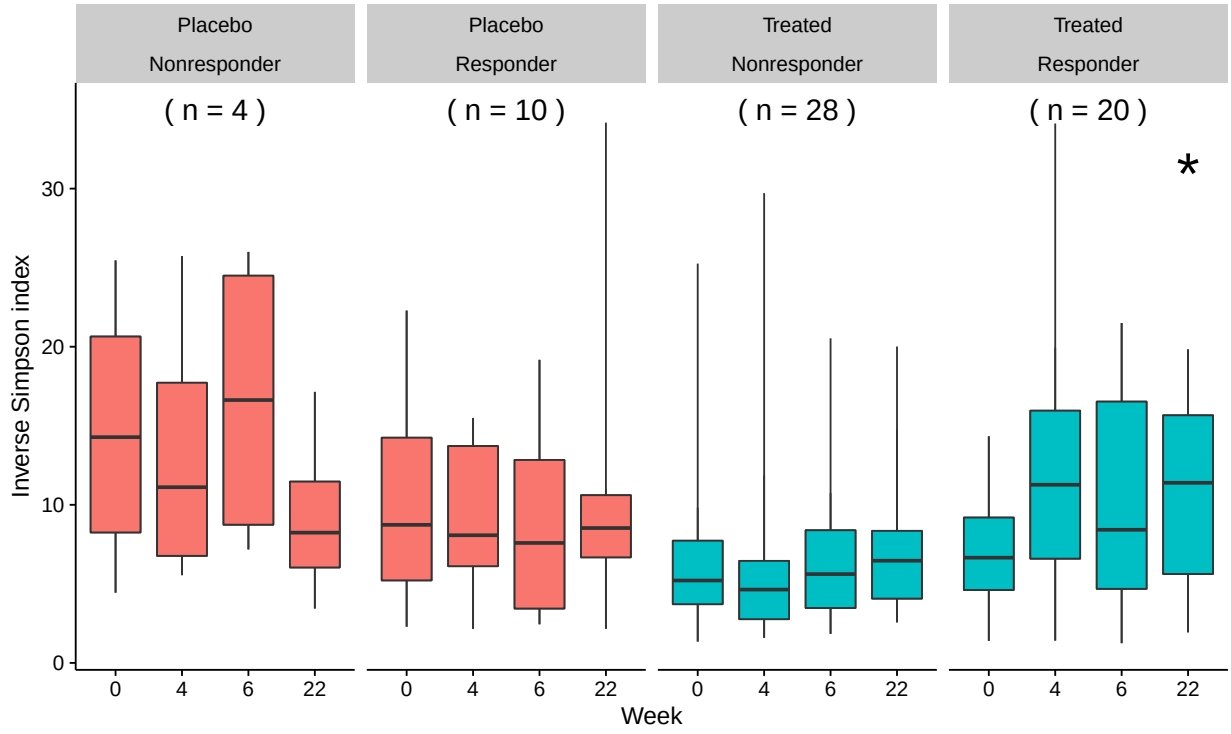
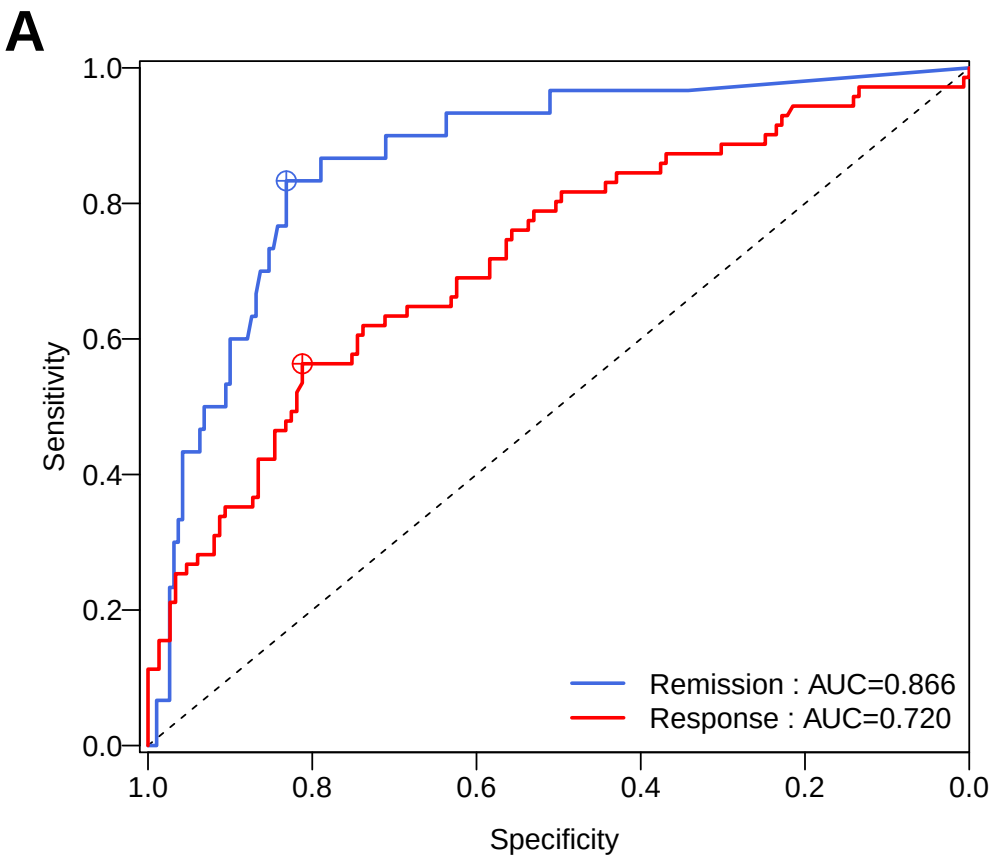
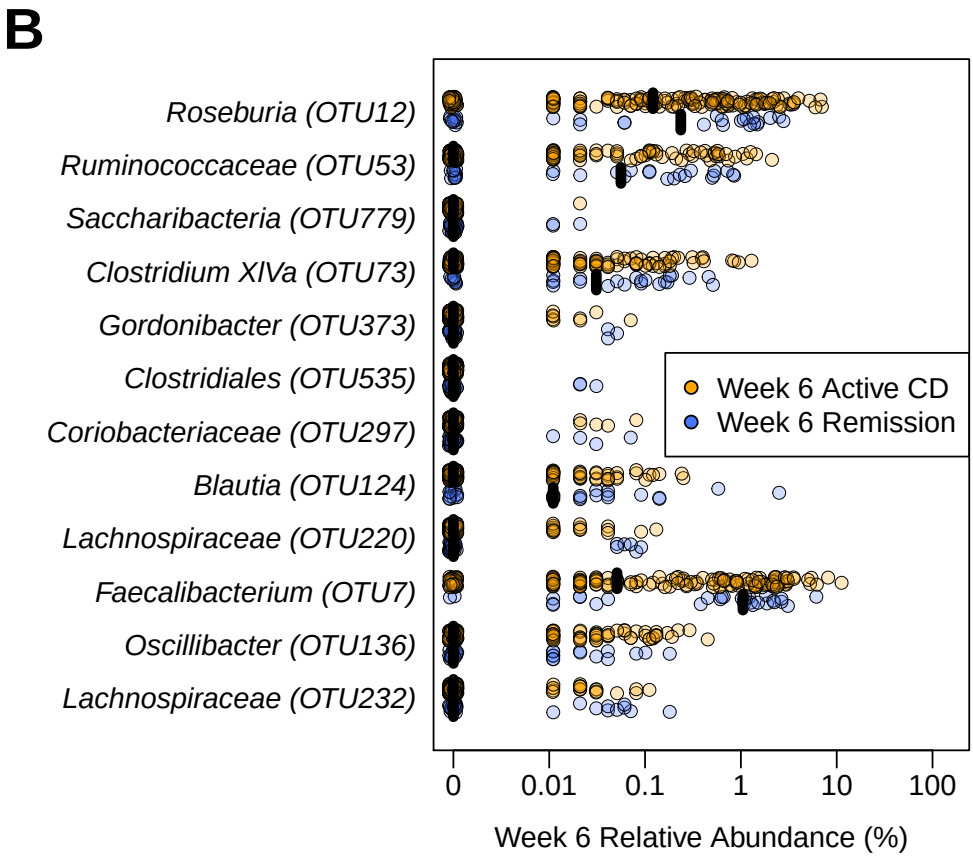


Figure 5: Classification of week 6 response or remission status using week 6 stool samples from subjects treated with UST (A) ROC curves for week 6 outcome based on the week 6 microbiota. (B) Predictive OTUs from week 6 stool for remission status at 6 weeks after induction, based on mean decrease in accuracy. Black bars represent the median relative abundance.


A
1.0
0.8
0.6
0.4
0.2
0.0
Sensitivity
1.0
0.8
0.6
0.4
0.2
0.0
Specificity
Remission : AUC=0.866
Response : AUC=0.720


B
Roseburia (OTU12)
Ruminococcaceae (OTU53)
Saccharibacteria (OTU779)
Clostridium XlVa (OTU73)
Gordonibacter (OTU373)
Clostridiales (OTU535)
Coriobacteriaceae (OTU297)
Blautia (OTU124)
Lachnospiraceae (OTU220)
Faecalibacterium (OTU7)
Oscillibacter (OTU136)
Lachnospiraceae (OTU232)
Week 6 Active CD
Week 6 Remission
0
0.01
0.1
1
10
100
Week 6 Relative Abundance (%)

398

References

1. Huang S, Li R, Zeng X, He T, Zhao H, Chang A, Bo C, Chen J, Yang F, Knight R, Liu J, Davis C, Xu J. 2014. Predictive modeling of gingivitis severity and susceptibility via oral microbiota. ISME J 8:1768–80.

2. Wang Y, Ames NP, Tun HM, Tosh SM, Jones PJ, Khafipour E. 2016. High molecular weight barley -glucan alters gut microbiota toward reduced cardiovascular disease risk. Front Microbiol 7.

3. Schubert AM, Sinani H, Schloss PD. 2015. Antibiotic-induced alterations of the murine gut microbiota and subsequent effects on colonization resistance against clostridium difficile. MBio 6:e00974.

4. Seekatz AM, Rao K, Santhosh K, Young VB. 2016. Dynamics of the fecal microbiome in patients with recurrent and nonrecurrent clostridium difficile infection. Genome Med 8.

5. Zackular JP, Rogers MA, Ruffin MT th, Schloss PD. 2014. The human gut microbiome as a screening tool for colorectal cancer. Cancer Prev Res (Phila) 7:1112–21.

6. Baxter NT, Ruffin MT th, Rogers MA, Schloss PD. 2016. Microbiota-based model improves the sensitivity of fecal immunochemical test for detecting colonic lesions. Genome Med 8:37.

7. Klatt NR, Cheu R, Birse K, Zevin AS, Perner M, Noel-Romas L, Grobler A, Westmacott G, Xie IY, Butler J, Mansoor L, McKinnon LR, Passmore JS, Abdool Karim Q, Abdool Karim SS, Burgener AD. 2017. Vaginal bacteria modify hiv tenofovir microbicide efficacy in african women. Science 356:938–945.

8. Haiser HJ, Gootenberg DB, Chatman K, Sirasani G, Balskus EP, Turnbaugh PJ. 2013. Predicting and manipulating cardiac drug inactivation by the human gut bacterium eggerthella lenta. Science 341:295–8.

9. Sivan A, Corrales L, Hubert N, Williams JB, Aquino-Michaels K, Earley ZM, Benyamin FW, Lei

YM, Jabri B, Alegre ML, Chang EB, Gajewski TF. 2015. Commensal bifidobacterium promotes antitumor immunity and facilitates anti-pd-l1 efficacy. Science 350:1084–9.

10. Vetizou M, Pitt JM, Daillere R, Lepage P, Waldschmitt N, Flament C, Rusakiewicz S, Routy B, Roberti MP, Duong CP, Poirier-Colame V, Roux A, Becharef S, Formenti S, Golden E, Cording S, Eberl G, Schlitzer A, Ginhoux F, Mani S, Yamazaki T, Jacquelot N, Enot DP, Berard M, Nigou J, Opolon P, Eggermont A, Woerther PL, Chachaty E, Chaput N, Robert C, Mateus C, Kroemer G, Raoult D, Boneca IG, Carbonnel F, Chamaillard M, Zitvogel L. 2015. Anticancer immunotherapy by ctla-4 blockade relies on the gut microbiota. Science 350:1079–84.

11. Gevers D, Kugathasan S, Denson LA, Vazquez-Baeza Y, Van Treuren W, Ren B, Schwager E, Knights D, Song SJ, Yassour M, Morgan XC, Kostic AD, Luo C, Gonzalez A, McDonald D, Haberman Y, Walters T, Baker S, Rosh J, Stephens M, Heyman M, Markowitz J, Baldassano R, Griffiths A, Sylvester F, Mack D, Kim S, Crandall W, Hyams J, Huttenhower C, Knight R, Xavier RJ. 2014. The treatment-naive microbiome in new-onset crohn's disease. Cell Host Microbe 15:382–92.

12. Wang F, Kaplan JL, Gold BD, Bhasin MK, Ward NL, Kellermayer R, Kirschner BS, Heyman MB, Dowd SE, Cox SB, Dogan H, Steven B, Ferry GD, Cohen SA, Baldassano RN, Moran CJ, Garnett EA, Drake L, Otu HH, Mirny LA, Libermann TA, Winter HS, Korolev KS. 2016. Detecting microbial dysbiosis associated with pediatric crohn disease despite the high variability of the gut microbiota. Cell Rep.

13. Ananthakrishnan AN, Luo C, Yajnik V, Khalili H, Garber JJ, Stevens BW, Cleland T, Xavier RJ. 2017. Gut microbiome function predicts response to anti-integrin biologic therapy in inflammatory bowel diseases. Cell Host Microbe 21:603–610.e3.

14. Shaw KA, Bertha M, Hofmekler T, Chopra P, Vatanen T, Srivatsa A, Prince J, Kumar A, Sauer C, Zwick ME, Satten GA, Kostic AD, Mulle JG, Xavier RJ, Kugathasan S. 2016. Dysbiosis, inflammation, and response to treatment: A longitudinal study of pediatric subjects with newly

448 diagnosed inflammatory bowel disease. Genome Med 8:75.

449 15. Ananthakrishnan AN. 2015. Epidemiology and risk factors for IBD. Nat Rev Gastroenterol
450 Hepatol 12:205–217.

451 16. Floyd DN, Langham S, Severac HC, Levesque BG. 2015. The economic and quality-of-life
452 burden of crohn's disease in europe and the united states, 2000 to 2013: A systematic review.
453 Dig Dis Sci 60:299–312.

454 17. Randall CW, Vizuete JA, Martinez N, Alvarez JJ, Garapati KV, Malakouti M, Taboada CM.
455 2015. From historical perspectives to modern therapy: A review of current and future biological
456 treatments for crohn's disease. Therap Adv Gastroenterol 8:143–59.

457 18. Wils P, Bouhnik Y, Michetti P, Flourie B, Brixi H, Bourrier A, Allez M, Duclos B, Grimaud
458 JC, Buisson A, Amiot A, Fumery M, Roblin X, Peyrin-Biroulet L, Filippi J, Bouguen G, Abitbol
459 V, Coffin B, Simon M, Laharie D, Pariente B. 2015. Subcutaneous ustekinumab provides clinical
460 benefit for two-thirds of patients with crohn's disease refractory to anti-tumor necrosis factor
461 agents. Clin Gastroenterol Hepatol.

462 19. Colombel JF, Reinisch W, Mantzaris GJ, Kornbluth A, Rutgeerts P, Tang KL, Oortwijn A,
463 Bevelander GS, Cornillie FJ, Sandborn WJ. 2015. Randomised clinical trial: Deep remission in
464 biologic and immunomodulator naive patients with crohn's disease - a SONIC post hoc analysis.
465 Aliment Pharmacol Ther 41:734–46.

466 20. Baert F, Moortgat L, Van Assche G, Caenepeel P, Vergauwe P, De Vos M, Stokkers P,
467 Hommes D, Rutgeerts P, Vermeire S, D'Haens G. 2010. Mucosal healing predicts sustained
468 clinical remission in patients with early-stage crohn's disease. Gastroenterology 138:463–8; quiz
469 e10–1.

470 21. Lichtenstein GR. 2010. Emerging prognostic markers to determine crohn's disease natural
471 history and improve management strategies: A review of recent literature. Gastroenterol Hepatol

(N Y) 6:99–107.

22. Chang S, Malter L, Hudesman D. 2015. Disease monitoring in inflammatory bowel disease. World J Gastroenterol 21:11246–59.

23. Boon GJ, Day AS, Mulder CJ, Gearry RB. 2015. Are faecal markers good indicators of mucosal healing in inflammatory bowel disease? World J Gastroenterol 21:11469–80.

24. Falvey JD, Hoskin T, Meijer B, Ashcroft A, Walmsley R, Day AS, Gearry RB. 2015. Disease activity assessment in ibd: Clinical indices and biomarkers fail to predict endoscopic remission. Inflamm Bowel Dis 21:824–31.

25. Sartor RB. 2006. Mechanisms of disease: Pathogenesis of crohn's disease and ulcerative colitis. Nat Clin Pract Gastroenterol Hepatol 3:390–407.

26. Wright EK, Kamm MA, Teo SM, Inouye M, Wagner J, Kirkwood CD. 2015. Recent advances in characterizing the gastrointestinal microbiome in crohn's disease: A systematic review. Inflamm Bowel Dis 21:1219–28.

27. Manichanh C, Rigottier-Gois L, Bonnaud E, Gloux K, Pelletier E, Frangeul L, Nalin R, Jarrin C, Chardon P, Marteau P, Roca J, Dore J. 2006. Reduced diversity of faecal microbiota in crohn's disease revealed by a metagenomic approach. Gut 55:205–11.

28. Hansen R, Russell RK, Reiff C, Louis P, McIntosh F, Berry SH, Mukhopadhya I, Bisset WM, Barclay AR, Bishop J, Flynn DM, McGrogan P, Loganathan S, Mahdi G, Flint HJ, El-Omar EM, Hold GL. 2012. Microbiota of de-novo pediatric IBD: Increased faecalibacterium prausnitzii and reduced bacterial diversity in crohn's but not in ulcerative colitis. Am J Gastroenterol 107:1913–22.

29. Haberman Y, Tickle TL, Dexheimer PJ, Kim MO, Tang D, Karns R, Baldassano RN, Noe JD, Rosh J, Markowitz J, Heyman MB, Griffiths AM, Crandall WV, Mack DR, Baker SS, Huttenhower C, Keljo DJ, Hyams JS, Kugathasan S, Walters TD, Aronow B, Xavier RJ, Gevers D, Denson LA. 2014. Pediatric crohn disease patients exhibit specific ileal transcriptome and microbiome

497 signature. J Clin Invest 124:3617–33.

498 30. Riol-Blanco L, Lazarevic V, Awasthi A, Mitsdoerffer M, Wilson BS, Croxford A, Waisman
499 A, Kuchroo VK, Glimcher LH, Oukka M. 2010. IL-23 receptor regulates unconventional il-17-
500 producing t cells that control infection1. J Immunol 184:1710–20.

501 31. Round JL, Mazmanian SK. 2009. The gut microbiome shapes intestinal immune responses
502 during health and disease. Nat Rev Immunol 9:313–23.

503 32. Eken A, Singh AK, Oukka M. 2014. INTERLEUKIN 23 in crohn'S disease. Inflamm Bowel
504 Dis 20:587–95.

505 33. Shih VFS, Cox J, Kljavin NM, Dengler HS, Reichelt M, Kumar P, Rangell L, Kolls JK, Diehl L,
506 Ouyang W, Ghilardi N. 2014. Homeostatic il-23 receptor signaling limits th17 response through
507 il-22–mediated containment of commensal microbiota. Proc Natl Acad Sci U S A 111:13942–7.

508 34. Tedjo DI, Smolinska A, Savelkoul PH, Masclee AA, Schooten FJ van, Pierik MJ, Penders J,
509 Jonkers DMAE. 2016. The fecal microbiota as a biomarker for disease activity in crohn's disease.
510 Scientific Reports, Published online: 13 October 2016; doi:101038/srep35216.

511 35. Sandborn WJ, Gasink C, Gao LL, Blank MA, Johanns J, Guzzo C, Sands BE, Hanauer
512 SB, Targan S, Rutgeerts P, Ghosh S, Villiers WJ de, Panaccione R, Greenberg G, Schreiber S,
513 Lichtiger S, Feagan BG. 2012. Ustekinumab induction and maintenance therapy in refractory
514 crohn's disease. N Engl J Med 367:1519–28.

515 36. Sandborn WJ, Feagan BG, Fedorak RN, Scherl E, Fleisher MR, Katz S, Johanns J, Blank M,
516 Rutgeerts P. 2008. A randomized trial of ustekinumab, a human interleukin-12/23 monoclonal
517 antibody, in patients with moderate-to-severe crohn's disease. Gastroenterology 135:1130–41.

518 37. Kopylov U, Afif W, Cohen A, Bitton A, Wild G, Bessissow T, Wyse J, Al-Taweel T, Szilagyi
519 A, Seidman E. 2014. Subcutaneous ustekinumab for the treatment of anti-TNF resistant crohn's

disease–the McGill experience. J Crohns Colitis 8:1516–22.

38. Peyrin-Biroulet L, Panes J, Sandborn WJ, Vermeire S, Danese S, Feagan BG, Colombel JF, Hanauer SB, Rycroft B. 2016. Defining disease severity in inflammatory bowel diseases: Current and future directions. Clin Gastroenterol Hepatol 14:348–354.e17.

39. Best WR, Becktel JM, Singleton JW, Kern J F. 1976. Development of a crohn's disease activity index. national cooperative crohn's disease study. Gastroenterology 70:439–44.

40. Calle ML, Urrea V, Boulesteix A-L, Malats N. 2011. AUC-RF: A new strategy for genomic profiling with random forest. Human Heredity 72:121–132.

41. Vogenberg FR. 2009. Predictive and prognostic models: Implications for healthcare decision-making in a modern recession. Am Health Drug Benefits 2:218–22.

42. Naftali T, Reshef L, Kovacs A, Porat R, Amir I, Konikoff FM, Gophna U. 2016. Distinct microbiotas are associated with ileum-restricted and colon-involving crohn's disease. Inflamm Bowel Dis 22:293–302.

43. Sartor RB, Wu GD. 2016. Roles for intestinal bacteria, viruses, and fungi in pathogenesis of inflammatory bowel diseases and therapeutic approaches. Gastroenterology.

44. Boon GJ, Day AS, Mulder CJ, Gearry RB. 2015. Are faecal markers good indicators of mucosal healing in inflammatory bowel disease? World J Gastroenterol 21:11469–80.

45. Chang S, Malter L, Hudesman D. 2015. Disease monitoring in inflammatory bowel disease. World J Gastroenterol 21:11246–59.

46. Papa E, Docktor M, Smillie C, Weber S, Preheim SP, Gevers D, Giannoukos G, Ciulla D, Tabbaa D, Ingram J, Schauer DB, Ward DV, Korzenik JR, Xavier RJ, Bousvaros A, Alm EJ. 2012. Non-invasive mapping of the gastrointestinal microbiota identifies children with inflammatory bowel disease. PLoS One 7:e39242.

47. Vandeputte D, Falony G, Vieira-Silva S, Tito RY, Joossens M, Raes J. 2016. Original

544 article: Stool consistency is strongly associated with gut microbiota richness and composition,
545 enterotypes and bacterial growth rates. Gut 65:57–62.

546 48. Huang EY, Inoue T, Leone VA, Dalal S, Touw K, Wang Y, Musch MW, Theriault B, Higuchi
547 K, Donovan S, Gilbert J, Chang EB. 2015. Using corticosteroids to reshape the gut microbiome:
548 Implications for inflammatory bowel diseases. Inflamm Bowel Dis 21:963–72.

549 49. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA,
550 Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF.
551 2009. Introducing mothur: Open-source, platform-independent, community-supported software
552 for describing and comparing microbial communities. Appl Environ Microbiol 75:7537–41.

553 50. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. 2013. Development of a
554 dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the
555 miseq illumina sequencing platform. Appl Environ Microbiol 79:5112–20.

556 51. Schloss PD, Gevers D, Westcott SL. 2011. Reducing the effects of PCR amplification and
557 sequencing artifacts on 16S rRNA-based studies. PLoS One 6:e27310.

558 52. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2013.
559 The silva ribosomal rna gene database project: Improved data processing and web-based tools.
560 Nucleic Acids Res 41:D590–6.

561 53. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. 2011. UCHIME improves sensitivity
562 and speed of chimera detection. Bioinformatics 27:2194–200.

563 54. Schloss PD, Westcott SL. 2011. Assessing and improving methods used in operational
564 taxonomic unit-based approaches for 16S rRNA gene sequence analysis. Appl Environ Microbiol
565 77:3219–26.

566 55. Wang Q, Garrity GM, Tiedje JM, Cole JR. 2007. Naive bayesian classifier for rapid assignment

of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol 73:5261–7.

56. R Core Team. 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

57. Sokal RR, Rohlf FJ. 1995. Biometry: The principles and practice of statistics in biological research, 3rd ed. Freeman, New York.

58. Magurran AE. 2004. Measuring biological diversity. Blackwell Pub., Malden, Ma.

59. Yue JC, Clayton MK. 2005. A similarity measure based on species proportions. Communications in Statistics-Theory and Methods 34:2123–2131.

60. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Szoecs E, Wagner H. 2016. Vegan: Community ecology package. r package version 2.4-1.

61. Friedman M. 1937. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. Journal of the American Statistical Association 32:675–701.

62. Giraudoux P. 2016. Pgirmess: Data analysis in ecology.

63. Urrea V, Calle M. 2012. AUCRF: Variable selection with random forest and the area under the curve.

64. Breiman L. 2001. Random forests. Machine Learning 45:5–32.

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

66. Wickham H. 2009. Ggplot2: Elegant graphics for data analysis. Springer-Verlag New York.

67. Wickham H, Francois R. 2016. Dplyr: A grammar of data manipulation.

68. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, Müller M. 2011. PROC:

590 An open-source package for r and s+ to analyze and compare roc curves. BMC Bioinformatics
591 12:77.

592 69. Xie Y. 2015. Dynamic documents with R and knitr, 2nd ed. Chapman; Hall/CRC, Boca
593 Raton, Florida.

594 70. Auguie B. 2016. GridExtra: Miscellaneous functions for "grid" graphics.

595 71. Wickham H, Chang W. 2016. Devtools: Tools to make developing r packages easier.

596 72. Boettiger C. 2015. Knitcitations: Citations for 'knitr' markdown files.

597 73. Wickham H. 2016. Scales: Scale functions for visualization.

598 74. Wickham H. 2017. Tidyr: Easily tidy data with 'spread()' and 'gather()' functions.

599 75. Harrell Jr FE, Charles Dupont, others. 2016. Hmisc: Harrell miscellaneous.

600 76. Wilke CO. 2016. Cowplot: Streamlined plot theme and plot annotations for 'ggplot2'.