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 the effect of immunomodulatory therapy on the microbiota and its role in response to therapy. This study explored associations between the fecal microbiota and 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 ($p = 0.012$). The observed baseline differences in fecal microbiota and changes due to therapeutic response support the potential for the microbiota as a response biomarker. (word count= 246/250, 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 identify patients more likely to respond 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 useful non-invasive biomarker for directing or monitoring the treatment of gastrointestinal diseases. (word count =98/150, 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 in African women (7), as well as cardiac drugs (8) and cancer treatments (9, 10) in murine models of disease. 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, including anti-integrin biologics (13, 14) and treament in pediatric IBD with anti-TNF-α or immunomodulators (15, 16). It remains to be determined, however, whether the composition of the fecal gut microbiota can predict and monitor response to biologic CD therapy directed at other targets, such as IL-23. Considering the involvement of the immune system and previous evidence for involvement of the microbiome, we hypothesize that response to anti-IL-23 CD therapy can be predicted using microbiome data.

CD is a global health concern causing large economic and health care impacts (17, 18). 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 (19, 20). Faster induction of remission following diagnosis reduces the risk of irreversible intestinal damage and disability (20–22). Ideally, clinicians would be able to determine personalized treatment options for CD patients at diagnosis that would result in faster achievement of remission (23). Therefore, recent research has been focused on identifying noninvasive biomarkers to monitor CD severity and predict therapeutic response (24–26).

The precise etiology of CD remains unknown, but host genetics, environmental exposure, and the gut microbiome appear to be involved (17, 27). Individuals with CD have reduced microbial diversity in their guts, compared to healthy individuals, with a lower relative abundance of *Firmicutes* and an increased relative abundance of *Enterobacteraciae* and *Bacteroides* (11, 28–31). 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 (19, 28, 32–35). 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 (36).

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 (20, 37–39). Given the potential impact of IL-23 on the microbiota (32–35), we hypothesized that response to UST could be 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 (19). 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 (37). 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, tested whether clinical responders had a microbiota that was distinct from non-responders, and determined 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 may be 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 (37). 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 100 points or remission. Remission was defined as a CDAI below 150 points. The CDAI is the standard instrument for evaluating clinical symptoms and disease activity in CD (40, 41). 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 (40, 41). 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). The number of subjects in each treatment group at the primary and secondary endpoints are summarized in Table 2 by their treatment outcome.

Association of baseline microbial signatures with treatment remission

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, 42). 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; 31 of which acheived remission (Table 2). 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 ($p = 0.841$).

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 = 0.033-3.70) 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 = 0.00-5.10), but it was also present prior to the start of treatment in every subject who was in remission 6 weeks after induction.

Association of baseline microbial signatures with treatment response

We also hypothesized that the composition of the baseline fecal microbiota could predict therapeutic response (CDAI decrease $\geq$ 100 points or remission) 6 weeks after induction. To test this hypothesis, we again used RF models to classify responders from non-responders 6 weeks after induction (Table 2). 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).

The microbiota model was significantly better able to predict response than the metadata alone (p = 0.017), whereas this was not true for the combined model (p = 0.069). Additionally, the combined model and the fecal microbiota model were not significantly different in their ability to predict response (p = 0.263). Optimal predictors were again determined based on their MDA in the ability of the model to classify response (Figure 2D). Also, our baseline combined model was significantly better at classifying remission compared to response (p = 0.036), whereas this was not true for the fecal microbiota model (p = 0.117).

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 to asses diversity measures (Table 2). 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 phenotypes

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 (36). 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 following UST therapy

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 responded to placebo (Figure 4). However, the median α-diversity of responders 22 weeks after UST induction significantly changed over time (p = 0.012) having increased from baseline

(median = 6.65, IQR = 4.60 - 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 (36). We again constructed RF classification models to distinguish between subjects by UST treatment outcome based on their fecal microbiota 6 weeks after induction (6, 42). 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 0.720 (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 identify patients more likely to achieve

remission following UST therapy, compared to clinical metadata alone, in this unique cohort. If this can be validated in future studies with independent cohorts, than it may lead to a clinically useful prognostic tool. 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 (43). 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 (28, 30, 44, 45). 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 (45). 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 achieve clinical remission rather than clinical response, 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 patient reported outcomes and more objectively quantifiable measures such as endoscopic verification of mucosal healing (21, 46), research is ongoing to discover less invasive and more quantifiable biomarkers (36, 47, 48).

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 lactoferrin 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 (36, 47, 48). 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, 36, 49). However, the CDAI sub score of weekly stool frequency may have driven these differences (Supplementary Table 1), which is consistent with previous studies (50). We also observed differences in the microbial community structure based on disease localization, which is consistent with a study by Naftali et al (44). Our study also found that corticosteroid use impacts the composition of the human fecal microbiota, which is consistent with observations in mouse models (51). 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 (36). 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 experimentally 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 in 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 (37) (Figure 1). Institutional review board approval was acquired at each participating study center and subjects provided written informed consent (37). 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 (37). 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 (48). Sequences were curated as described previously using the mothur software package (v.1.34.4) (52, 53). Briefly, we curated the sequences to reduce sequencing and PCR errors (54), aligned the resulting sequences to the SILVA 16S rRNA sequence database (55), and used UCHIME to remove any chimeric sequences (56). Sequences were clustered into operational

taxonomic units (OTU) at a 97% similarity cutoff using the average neighbor algorithm (57). 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 (58).

Following sequence curation using the mothur software package (52), 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 (59). 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 (60, 61). 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 (62). 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) (63). 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 (64). 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) (65). 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) (66), to identify phylotypes or clinical variables that distinguish between various treatment and response groups, as well as to predict or determine response outcome (67). 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 (68). Other R packages used in our analysis included ggplot2 v.2.2.1 (69), dplyr v.0.5.0 (70), pROC v.1.9.1 (71), knitr v.1.15.1 (72), gridExtra v.2.2.1 (73), devtools v.1.12.0 (74), knitcitations v.1.0.7 (75), scales v.0.4.1 (76), tidyr v.0.6.1 (77), Hmisc v.4.0.2 (78), and cowplot v.0.7.0 (79). 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

Table 2: Summary of subjects in each treatment group by endpoint and outcome

Clinical Variable	Treated (n)	Placebo (n)
Week 6 Response (No, Yes)	156, 76	48, 26
Week 6 Remission (No, Yes)	201, 31	62, 12
Week 22 Response (No, Yes)	77, 43	14, 11
Week 22 Remission (No, Yes)	96, 24	18, 7

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.

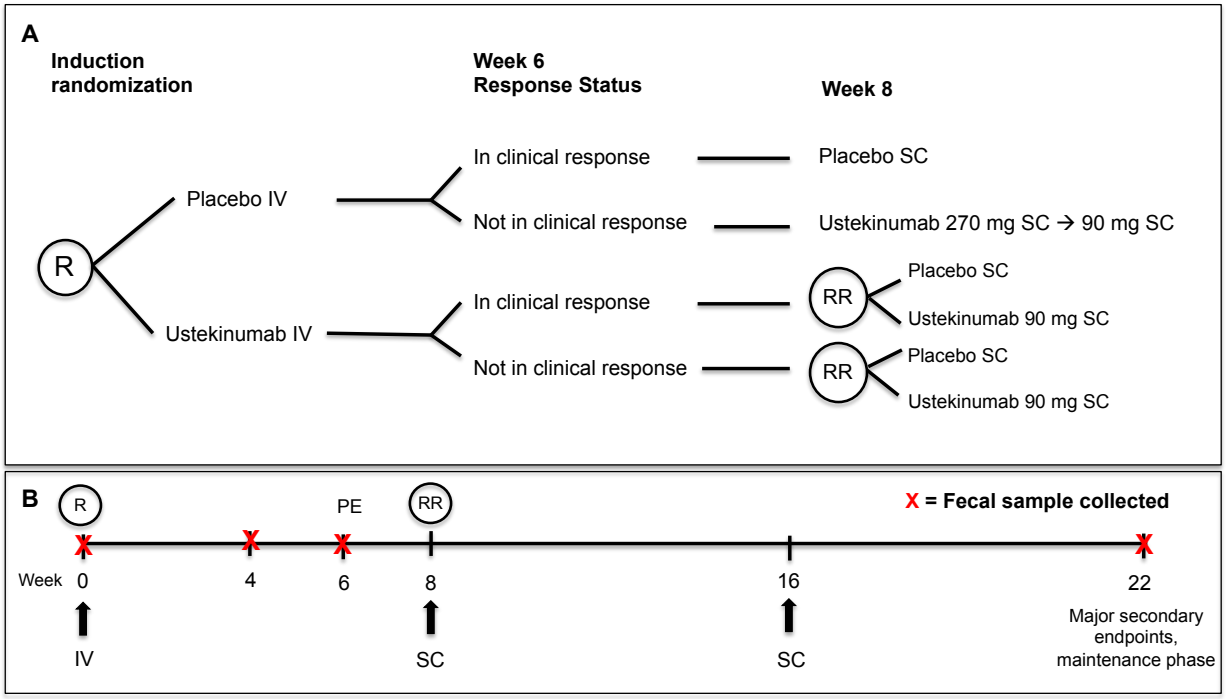
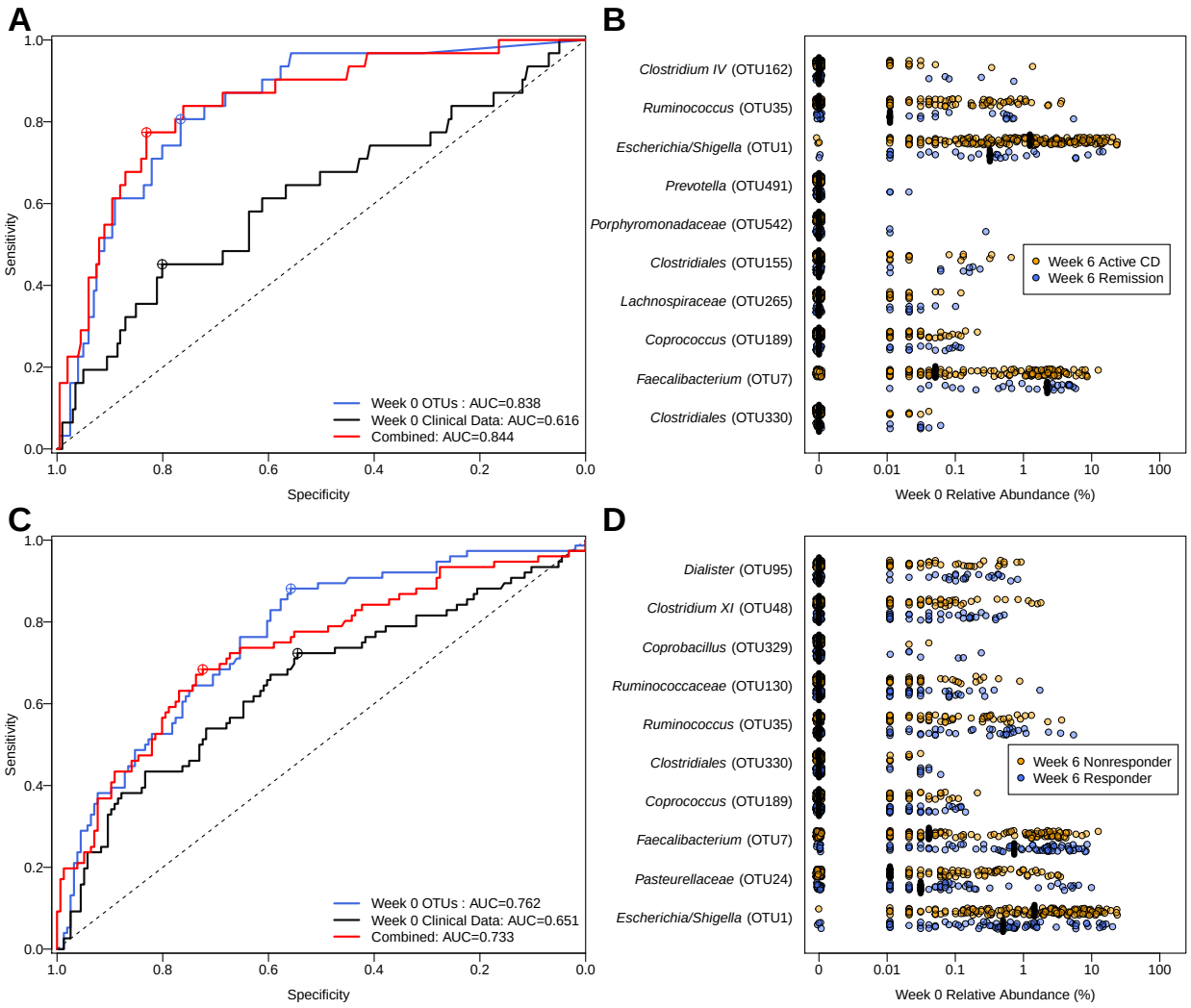


Figure 2: Prediction of week 6 treatment outcome in subjects treated with UST, using baseline samples Receiver operating characteristic (ROC) curves for (A) response and (C) remission using microbiota data (blue), clinical metadata (black), and a combined model (red). Top predictive OTUs for the microbiota 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).

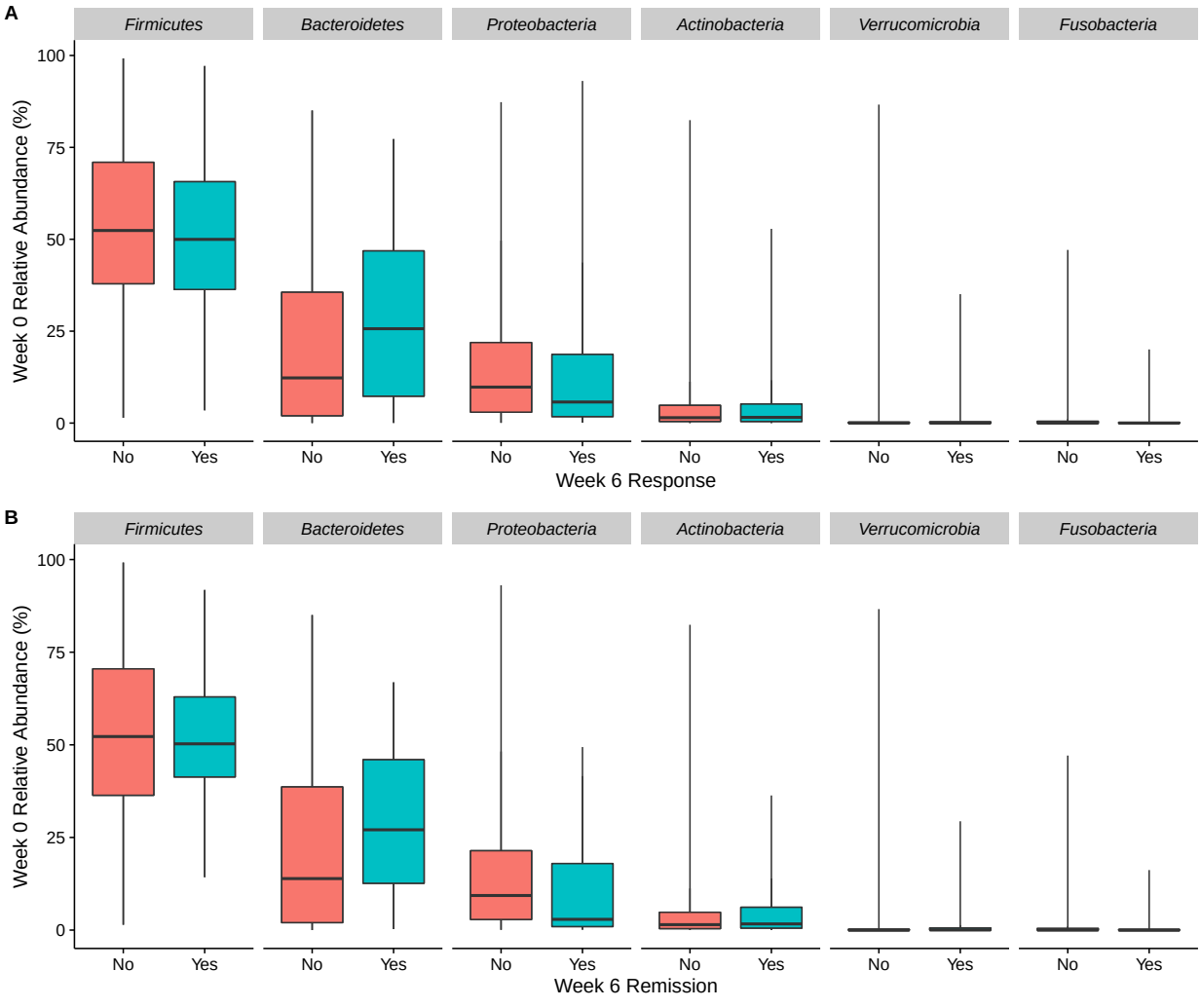


Figure 3: Differential taxa in baseline stool samples from subjects treated with UST, based on week six remission status The baseline relative abundance of each OTU was compared between subjects in remission and those with active CD 6 weeks after induction using a Wilcoxon rank sum test followed by a 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.

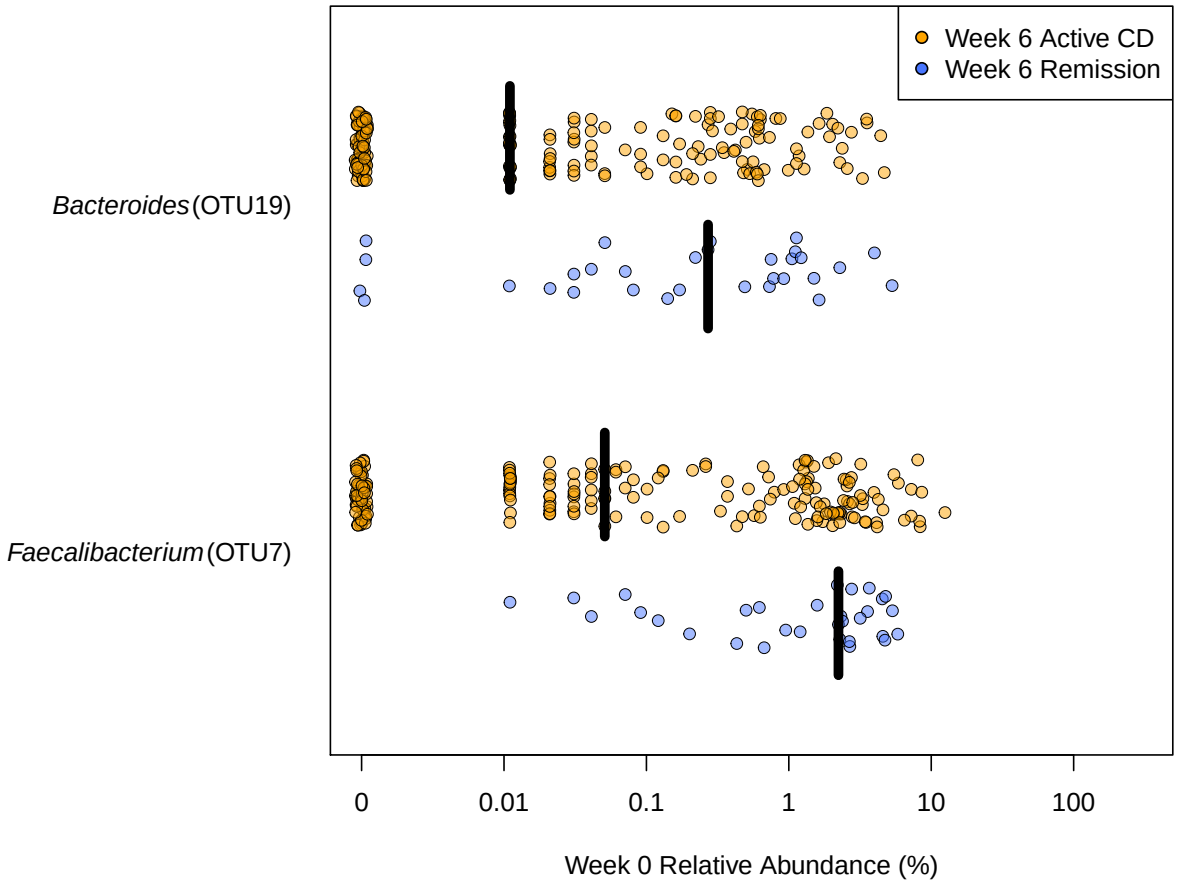


Figure 4: Change in alpha diversity over time by induction treatment and week 22 response status. The α-diversity of 48 subjects induced and maintained with UST and 14 subjects induced and maintained with placebo was assessed at each time point. Friedman test were performed within each treatment and 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 ($p < 0.05$).

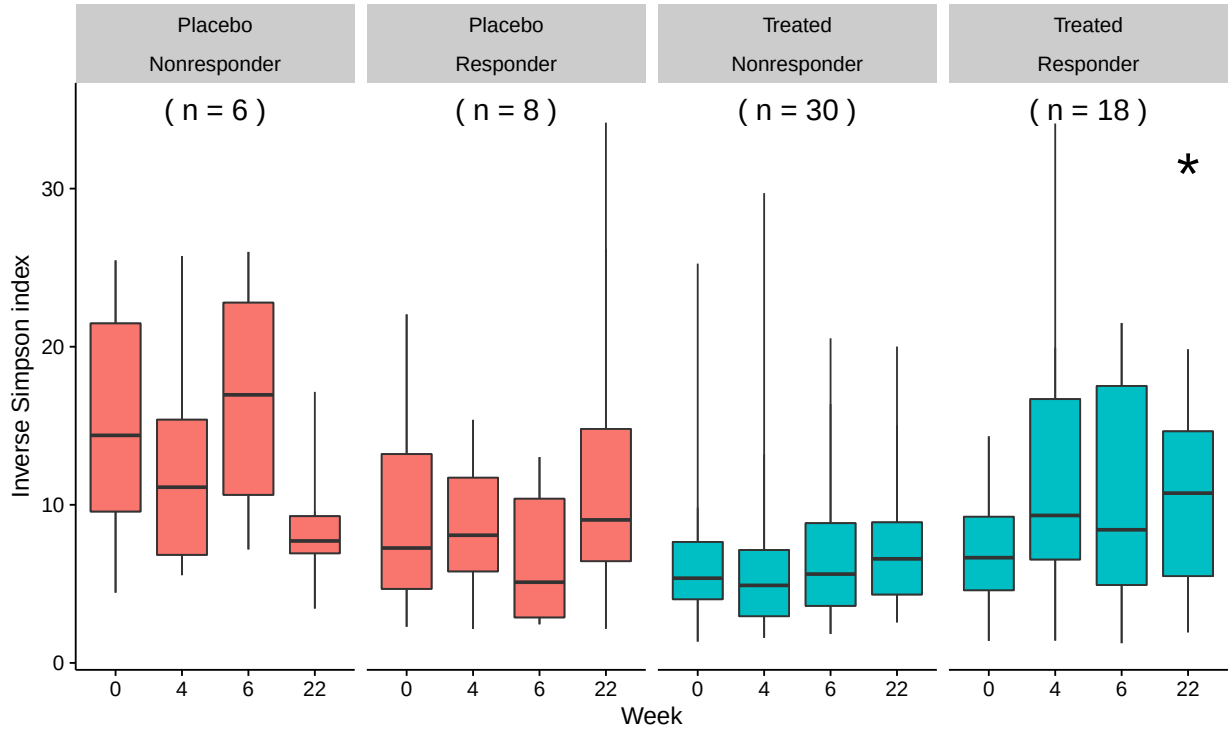
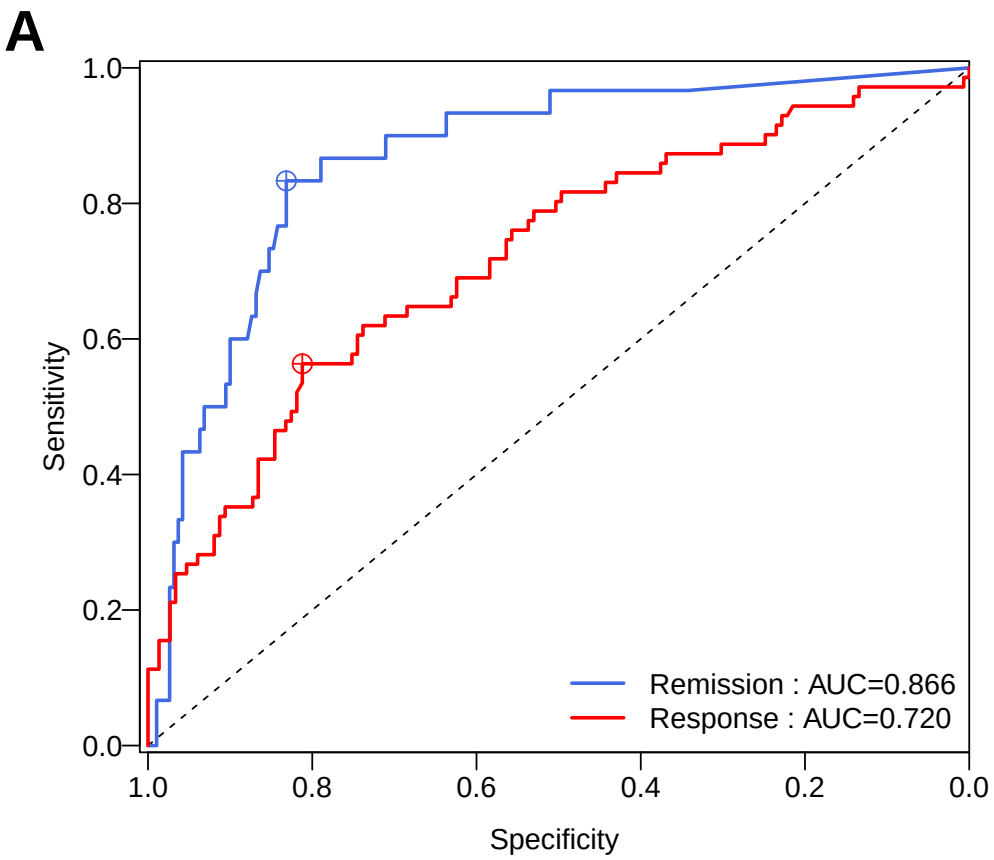
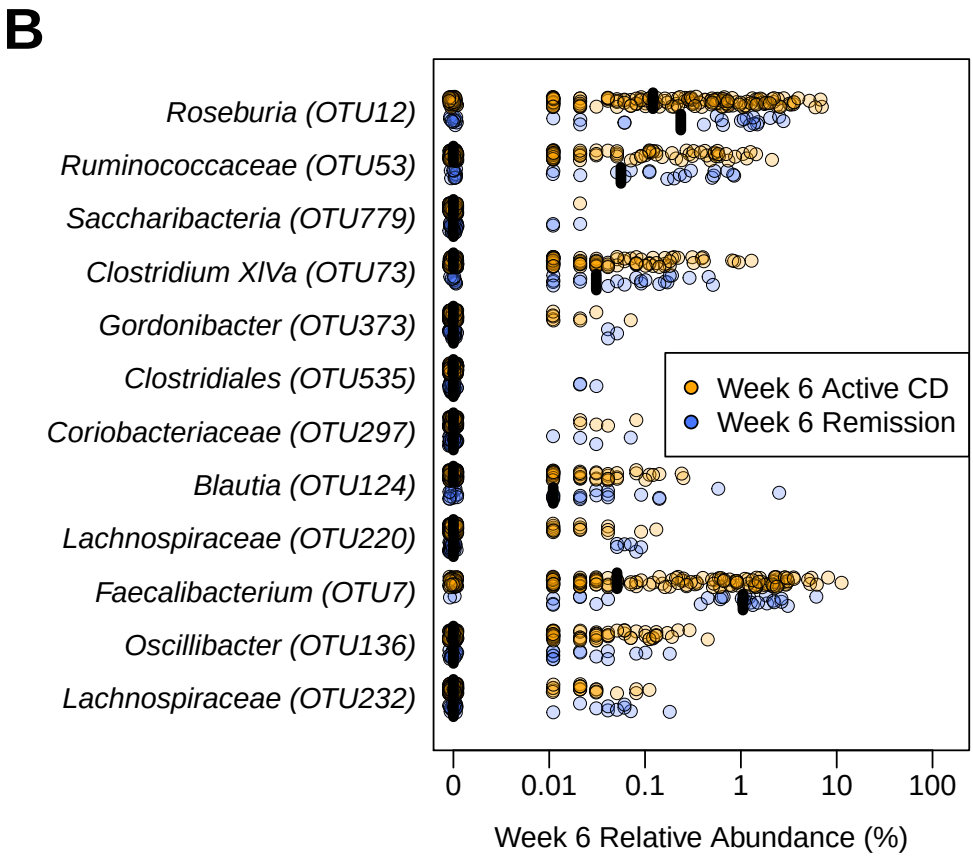


Figure 5: Classification of week 6 response or remission status using week 6 stool samples from subjects treated with UST (A) ROC curves for week 6 outcome based on the week 6 microbiota. (B) Predictive OTUs from week 6 stool for remission status at 6 weeks after induction, based on mean decrease in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

411

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

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

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

444 11. Gevers D, Kugathasan S, Denson LA, Vazquez-Baeza Y, Van Treuren W, Ren B, Schwager
445 E, Knights D, Song SJ, Yassour M, Morgan XC, Kostic AD, Luo C, Gonzalez A, McDonald D,
446 Haberman Y, Walters T, Baker S, Rosh J, Stephens M, Heyman M, Markowitz J, Baldassano R,
447 Griffiths A, Sylvester F, Mack D, Kim S, Crandall W, Hyams J, Huttenhower C, Knight R, Xavier
448 RJ. 2014. The treatment-naive microbiome in new-onset crohn's disease. Cell Host Microbe
449 15:382–92.

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

455 13. Tew GW, Hackney JA, Gibbons D, Lamb CA, Luca D, Egen JG, Diehl L, Eastham Anderson
456 J, Vermeire S, Mansfield JC, Feagan BG, Panes J, Baumgart DC, Schreiber S, Dotan I, Sandborn
457 WJ, Kirby JA, Irving PM, De Hertogh G, Van Assche GA, Rutgeerts P, O'Byrne S, Hayday A, Keir
458 ME. 2016. Association between response to etrolizumab and expression of integrin alphaE and
459 granzyme a in colon biopsies of patients with ulcerative colitis. Gastroenterology 150:477–87.e9.

460 14. Ananthakrishnan AN, Luo C, Yajnik V, Khalili H, Garber JJ, Stevens BW, Cleland T, Xavier
461 RJ. 2017. Gut microbiome function predicts response to anti-integrin biologic therapy in inflam-

matory bowel diseases. Cell Host Microbe 21:603–610.e3.

15. Kolho KL, Korpela K, Jaakkola T, Pichai MV, Zoetendal EG, Salonen A, Vos WM de. 2015. Fecal microbiota in pediatric inflammatory bowel disease and its relation to inflammation. Am J Gastroenterol 110:921–30.

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

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

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

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

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

21. Colombel JF, Reinisch W, Mantzaris GJ, Kornbluth A, Rutgeerts P, Tang KL, Oortwijn A, Bevelander GS, Cornillie FJ, Sandborn WJ. 2015. Randomised clinical trial: Deep remission in biologic and immunomodulator naive patients with crohn's disease - a SONIC post hoc analysis.

Aliment Pharmacol Ther 41:734–46.

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

23. Lichtenstein GR. 2010. Emerging prognostic markers to determine crohn's disease natural history and improve management strategies: A review of recent literature. Gastroenterol Hepatol (N Y) 6:99–107.

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

25. 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.

26. 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.

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

28. 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.

29. 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.

30. Hansen R, Russell RK, Reiff C, Louis P, McIntosh F, Berry SH, Mukhopadhya I, Bisset WM,

510 Barclay AR, Bishop J, Flynn DM, McGrogan P, Loganathan S, Mahdi G, Flint HJ, El-Omar EM,
511 Hold GL. 2012. Microbiota of de-novo pediatric IBD: Increased faecalibacterium prausnitzii and
512 reduced bacterial diversity in crohn's but not in ulcerative colitis. Am J Gastroenterol 107:1913–
513 22.

514 31. Haberman Y, Tickle TL, Dexheimer PJ, Kim MO, Tang D, Karns R, Baldassano RN, Noe JD,
515 Rosh J, Markowitz J, Heyman MB, Griffiths AM, Crandall WV, Mack DR, Baker SS, Huttenhower
516 C, Keljo DJ, Hyams JS, Kugathasan S, Walters TD, Aronow B, Xavier RJ, Gevers D, Denson
517 LA. 2014. Pediatric crohn disease patients exhibit specific ileal transcriptome and microbiome
518 signature. J Clin Invest 124:3617–33.

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

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

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

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

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

532 37. Sandborn WJ, Gasink C, Gao LL, Blank MA, Johanns J, Guzzo C, Sands BE, Hanauer
533 SB, Targan S, Rutgeerts P, Ghosh S, Villiers WJ de, Panaccione R, Greenberg G, Schreiber S,
534 Lichtiger S, Feagan BG. 2012. Ustekinumab induction and maintenance therapy in refractory

crohn's disease. N Engl J Med 367:1519–28.

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

39. Kopylov U, Afif W, Cohen A, Bitton A, Wild G, Bessissow T, Wyse J, Al-Taweel T, Szilagyi 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.

40. 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.

41. 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.

42. 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.

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

44. 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.

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

46. Williet N, Sandborn WJ, Peyrin-Biroulet L. 2014. Patient-reported outcomes as primary end points in clinical trials of inflammatory bowel disease. Clin Gastroenterol Hepatol 12:1246–56.e6.

47. Boon GJ, Day AS, Mulder CJ, Gearry RB. 2015. Are faecal markers good indicators of

559 mucosal healing in inflammatory bowel disease? World J Gastroenterol 21:11469–80.

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

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

566 50. Vandeputte D, Falony G, Vieira-Silva S, Tito RY, Joossens M, Raes J. 2016. Original
567 article: Stool consistency is strongly associated with gut microbiota richness and composition,
568 enterotypes and bacterial growth rates. Gut 65:57–62.

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

572 52. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA,
573 Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF.
574 2009. Introducing mothur: Open-source, platform-independent, community-supported software
575 for describing and comparing microbial communities. Appl Environ Microbiol 75:7537–41.

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

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

581 55. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2013.
582 The silva ribosomal rna gene database project: Improved data processing and web-based tools.

583 Nucleic Acids Res 41:D590–6.

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

586 57. Schloss PD, Westcott SL. 2011. Assessing and improving methods used in operational
587 taxonomic unit-based approaches for 16S rRNA gene sequence analysis. Appl Environ Microbiol
588 77:3219–26.

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

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

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

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

596 62. Yue JC, Clayton MK. 2005. A similarity measure based on species proportions. Communica-
597 tions in Statistics-Theory and Methods 34:2123–2131.

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

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

603 65. Giraudoux P. 2016. Pgirmess: Data analysis in ecology.

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

the curve.

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

68. 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.

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

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

71. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, Müller M. 2011. PROC: An open-source package for r and s+ to analyze and compare roc curves. BMC Bioinformatics 12:77.

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

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

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

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

76. Wickham H. 2016. Scales: Scale functions for visualization.

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

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

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