



Published in final edited form as:

Cell Host Microbe. 2021 August 11; 29(8): 1294–1304.e4. doi:10.1016/j.chom.2021.06.019.

Multi-omics reveal microbial determinants impacting responses to biologic therapies in Inflammatory Bowel Disease

Jonathan Wei Jie Lee^{1,2,3,4}, Damian Plichta¹, Larson Hogstrom^{1,†}, Nynke Z. Borren^{5,‡}, Helena Lau^{1,5}, Sara M. Gregory⁵, William Tan⁵, Hamed Khalili⁵, Clary Clish¹, Hera Vlamakis¹, Ramnik J. Xavier^{1,6,7,8,**}, Ashwin N Ananthakrishnan^{5,**}

¹Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA

²Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, 1E Kent Ridge Road, Singapore 119228, Singapore

³Division of Gastroenterology and Hepatology, National University Health System, Singapore University Medical Center, 5 Lower Kent Ridge Rd, Singapore 119074, Singapore

⁴NUS Synthetic Biology for Clinical and Technological Innovation (SynCTI), 14 Medical Drive, MD6-Centre for Translational Medicine, Singapore 117599, Singapore

⁵Division of Gastroenterology, Massachusetts General Hospital, Boston, MA 02114, USA

⁶Center for Microbiome Informatics and Therapeutics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

⁷Center for Computational and Integrative Biology and Department of Molecular Biology, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114, USA

⁸Lead Contact

SUMMARY

The intestinal microbiome is a key determinant of responses to biologic therapy in inflammatory bowel disease (IBD). However, diverse therapeutics and variable responses among IBD patients have posed challenges in predicting clinical therapeutic success. In this prospective study, we profiled baseline stool and blood in patients with moderate-to-severe Crohn's disease or ulcerative

**Co-corresponding and senior authors: xavier@molbio.mgh.harvard.edu, aananthakrishnan@mgh.harvard.edu.

†Present address: Thrive Earlier Detection Corp, Cambridge, MA, USA

‡Present address: University of Groningen, Groningen, The Netherlands

AUTHOR CONTRIBUTIONS

A.N.A., R.J.X. conceived and designed the study. A.N.A., N.B., H.K., R.J.X., W.L., H.L., S.M.G., enrolled the patients and conducted sampling and clinical measures. C.C., generated the metabolomics data. J.W.J.L., L.H., analyzed sequencing data from metagenomics, metabolomics and proteomics. D.R.P. assembled and analyzed metagenomes for *Bai* enzyme-encoding species. J.W.J.L., D.R.P., A.N.A., H.V., R.J.X., discussed results and wrote the paper, and all authors read, discussed and approved the final manuscript.

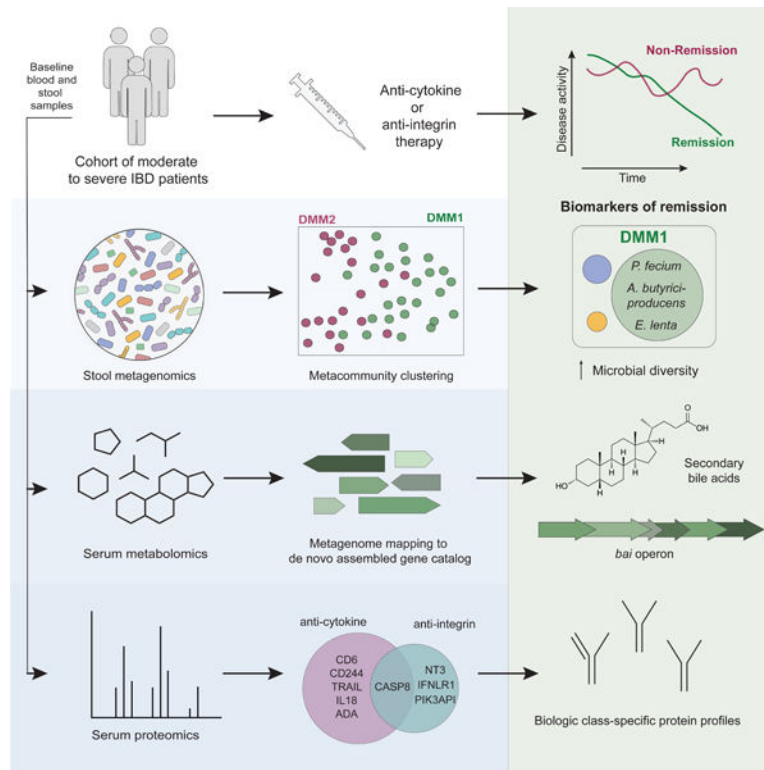
Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

DECLARATION OF INTERESTS

R.J.X. is a consultant to Novartis and co-founder of Celsius Therapeutics and Jnana Therapeutics. A.N.A. serves as a scientific advisor to Abbvie, Sun Pharma, Ikena Therapeutics and Gilead. J.L.W.J. is co-founder of AMILI and serves as a member of the scientific advisory board.

colitis initiating anti-cytokine therapy (anti-TNF or -IL12/23) or anti-integrin therapy. Patients were assessed at 14 weeks for clinical remission and 52 weeks for clinical and endoscopic remission. Baseline microbial richness indicated preferential responses to anti-cytokine therapy and correlated with abundance of microbial species capable of 7 α / β -dehydroxylation of primary to secondary bile acids. Serum signatures of immune-proteins reflecting microbial diversity identified patients more likely to achieve remission with anti-cytokine therapy. Remission-associated multi-omic profiles were unique to each therapeutic class. These profiles may facilitate *a priori* determination of optimal therapeutics for patients and serve as targets for newer therapies.

Graphical Abstract



eTOC BLURB

Biomarkers predicting responses and guiding stratified therapeutic approaches in Inflammatory Bowel Diseases (IBD) remains an unmet medical need. Lee et al. identified fecal metagenomic, serum metabolomic and proteomic markers that predict differential response to either anti-cytokine or anti-integrin therapy in IBD.

INTRODUCTION

Inflammatory bowel diseases (IBD; Crohn's disease (CD), ulcerative colitis (UC)) affect nearly 3 million individuals in the United States (Ananthakrishnan et al., 2020; Graham and Xavier, 2020). The availability of therapies with diverse mechanisms of action inhibiting leukocyte trafficking (anti-integrins) or blocking the effect of inflammatory cytokines (anti-

tumor necrosis factor, anti-interleukin (IL) 12/23) (Danese et al., 2015) have revolutionized our ability to achieve clinical remission and endoscopic healing, thereby preventing disease-related morbidity. However, despite an initial response, only a small proportion of patients achieve sustained remission. In addition, a critically important limitation of the existing therapeutic paradigm is the need for sequential trials of different therapeutic classes upon failure of previous therapy (Plichta et al., 2019). The inability to *a priori* define the medication most likely to benefit a patient introduces a potential delay in achieving remission, reduces the rate of response, and leads to protracted morbidity. Thus, understanding the determinants of response to one or more therapeutic classes is of paramount scientific and clinical importance. In observational cohorts and randomized controlled trials, clinical parameters alone have insufficient predictive value in determining response to a therapeutic class and there has been no robust development of models that may predict differential response to different mechanisms of action.

The intestinal microbiome is central to the pathogenesis of CD and UC (Gevers et al., 2014; Lloyd-Price et al., 2019; Schirmer et al., 2018). Diversion of the fecal stream with an ileostomy effectively ameliorates downstream inflammation in the colon (Rutgeerts et al., 1991). A less diverse microbiome, greater abundance of pro-inflammatory species such as *Escherichia coli* (Rhodes, 2007) and *Ruminococcus gnavus* (Hall et al., 2017) and depletion of species with protective mechanisms such as short chain fatty acid production is associated with intestinal inflammation in IBD. In prior work (Ananthakrishnan et al., 2017), we demonstrated that a machine-learning algorithm incorporating functional and taxonomic microbial data from stool performed accurately in predicting initial response to vedolizumab therapy. However, this prior study did not incorporate metabolomic profiling which provides a window into the functional impact of the gut microbiome, and was limited in its duration of follow-up to examine clinical and endoscopic outcomes at 1 year after treatment. Similarly, others have demonstrated that the baseline fecal microbiome may predict response to anti-TNF or anti-IL12/23 therapy individually (Aden et al., 2019; Ananthakrishnan et al., 2017; Estevinho et al., 2020). However, all these studies were limited by small numbers of included patients and examination primarily of short-term improvement with one therapeutic class, precluding definition of whether microbial profiles differentially predict response based on the therapeutic mechanism of action. Here, we profile the gut metagenome of patients with moderate-to-severe CD or UC initiating anti-TNF or anti-IL12/23 (together termed anti-cytokine), or anti-integrin therapy to define microbial features that predict response to one or more therapeutic classes. By pairing metagenomic sequencing with serum proteomic and metabolomic profiles, we define the potential functional mechanisms of the impact of the microbiome on the outcome.

RESULTS

To study the microbial determinants of clinical outcomes, we identified eligible IBD patients initiating biologic therapy enrolled in the Prospective Registry in IBD Study at Massachusetts General Hospital (MGH) (PRISM) (Franzosa et al., 2019). The study cohort (n=185; 108 CD, 77 UC) consisted of 79 (42.7%) patients with moderate-to-severe IBD initiating anti-TNF, 21 (11.4%) initiating ustekinumab (anti-IL12/23) and 85 (45.9%) initiating vedolizumab (anti-integrin) therapy (Figure 1, Table S1). Anti-TNF and anti-

IL12/23 were grouped together as anti-cytokine therapy. Stool and blood samples were collected at baseline, and each stool sample was subjected to metagenomic sequencing followed by profiling of the microbial community taxonomic composition and functional potential. Blood samples were analyzed for untargeted metabolomics and serum proteomics, to capture metabolomic and immune signaling profiles associated with putative microbial features associated with response to different therapies. By the end of 1 year of follow-up, 91 (49.2%) achieved clinical remission status by 14 weeks and an additional 22 (11.9%) patients achieved remission at 52 weeks (Table S2). Among the 86 patients who underwent colonoscopy at 1 year, 41 (47.7%) achieved endoscopic remission.

Baseline microbial metacommunity differentially predicts response to anti-cytokine therapy

Consistent with prior smaller studies, patients treated with anti-cytokine therapy who achieved clinical remission at 14 weeks ($p < 0.1$) and at one year ($p < 0.01$) had significantly higher indices of microbial species richness at baseline (Figure S1A–C), and likewise for Shannon diversity although the latter was not statistically significant. The indices of alpha diversity (both richness and Shannon) positively correlated ($p < 0.05$) with a diagnosis of UC, concomitant use of steroids, immunomodulators, and prior TNF exposure, and were negatively correlated with concomitant use of antibiotics (Figure S1D,E).

Using Dirichlet multinomial mixture (DMM) models to partition microbial community profiles into a finite number of clusters using Laplace approximation, we identified two distinct metacommunities, designated ‘DMM group 1’ and ‘DMM group 2’ (Figure 2A). The optimal number of clusters determined was congruent with results from a partitioning around medoid (PAM)-based clustering approach (see Figure S1F). Metacommunity DMM group 1 was represented predominantly by microbial species of the Firmicutes phyla including *Faecalibacterium prausnitzii* and *Ruminococcus bromii*. DMM group 1 also had a more diverse community profile (Figure 2B). In contrast, DMM group 2 was represented by *Bacteroides ovatus*, *B. thetaiotaomicron*, and *Veillonella parvula* among others (Figure 2C). Baseline community membership was an important determinant of clinical outcomes depending on the therapeutic mechanism of action (Figure 2D). Patients on anti-cytokine therapy who subsequently achieved clinical remission at 14 weeks, at one year, and endoscopic remission at 1 year, had higher baseline microbial richness. For patients initiating anti-cytokine therapy, the rates of clinical remission at week 14 (73.3% vs. 41.2%, $p = 0.02$), week 52 (66.7% vs. 36.4%, $p < 0.01$) and endoscopic remission (65% vs. 37.5%, $p = 0.09$) were much higher in those belong to DMM group 1 at baseline when compared to group 2.

The microbiome profiles amongst IBD patients may differ significantly, depending on IBD subtype (i.e. CD and UC) and initial disease severity (Lloyd-Price et al., 2019; Schirmer et al., 2018). We thereby performed additional sub-analyses, by further sub-stratifying the proportion of week 14 responders to either anti-cytokine therapy or vedolizumab therapy, according to subtype of IBD or baseline disease severity and their corresponding baseline metagenomic DMM group. We also used a cut off of 8 with the Harvey Bradshaw index for patients with CD, and cut off of 5 with the simple clinical colitis activity index for

patients with UC, to define moderate-severe IBD activity (Peyrin-Biroulet et al., 2016), to investigate preferential response to either anti-cytokine or anti-integrin therapy. Of note, the preferential response to anti-cytokine therapy to anti-integrin therapy amongst patients with baseline DMM group 1 metacommunities is independent of the IBD subtype (CD; 77.3% vs 45.4%, UC; 57.1% vs 39.3%, Figure S1G) or baseline disease severity (Mild; 83.3% vs 62.5%, Moderate-Severe; 64.7% vs 41.7%, Figure S1H).

Taxonomic and functional metagenomic features associated with remission

To test association with clinical outcomes after exposure to either therapeutic class, whilst controlling for other covariates (such as age, subtype of IBD, duration of disease, and medication use), we applied a multivariable linear model to each microbial taxonomic feature, present in at least 10 samples with 0.1% relative abundance. The abundances of nine microbial species at baseline ($qval < 0.2$) were linked to early clinical remission amongst patients treated with anti-cytokine and three microbial species for anti-integrin therapy (Figure 2E, Table S3). Several of these species including *Phascolarctobacterium faecium*, *Agathobaculum butyriciproducens*, *Clostridium citroneae* have previously demonstrated anti-inflammatory effects, primarily through altering the fecal short-chain fatty acid production (Watanabe et al., 2012). Interestingly, certain species specifically associated with remission either for anti-cytokine therapy (example *C. citroneae*, *A. butyriciproducens*) or anti-integrin alone (*B. stercoris*) while others demonstrated similar effects for both therapeutic mechanisms (*B. caccae* and *B. ovatus*). Overall, the association between baseline microbiome and remission was stronger for week 14 outcomes for both therapeutic mechanisms than with week 52 clinical or endoscopic remission, suggesting the clinical impact of the microbiome is strongest on more proximate events.

To understand the functional consequences of microbial composition changes leading to remission, we profiled the gene families in all metagenomes using HUMAnN3, and then summed their abundances according to the Enzyme Commission (EC) number annotations. We then applied the above described linear modelling approach to enzyme abundance data, revealing 120 enzymes, from six enzyme classes, that were mostly differentially abundant ($qval < 0.2$) in baseline samples from patients receiving anti-cytokine who achieved early remission compared to non-remitters (Table S3). When examining species-level functional attribution data, 6.8% (8/120) of the enzymes were explained by a single species contributing >50% of the enzyme copies in >50% of the samples (hereafter single-species dominance). *Eggerthella lenta* alone dominated 5 of the 8 enzymes, 3 of which (EC 1.1.1.391, EC 3.1.2.26, EC 1.1.1.52) are associated with secondary bile acid biosynthesis. The remaining 112 enzymes that were not associated with single-species dominance were mostly enriched in non-remitter subjects, compared to those with early remission. Two-fifths of these enzymes (28/70) belonged to class EC3 (hydrolase activity; mainly EC 3.2, acting on glycosyl bonds) (Figure S2). Glycoside hydrolases are involved in the breakdown and utilization of mucin-derived glycans and the loss of particular glycans, such as complex O-glycans, can result in spontaneous colitis (Fu et al., 2011).

Serum secondary bile acids predict clinical remission

We conducted metabolome analysis on baseline blood samples from patients undergoing anti-cytokine therapy, using four liquid chromatography tandem mass spectrometry (LC-MS) methods previously described (Franzosa et al., 2019). A total of 541 unique features were matched against reference data generated from an in-house compound library, and binned into metabolite classes as defined by the Human Metabolome Database (HMDB) (Figure S3A). Using the same multivariable linear regression model described above, a total of 27 putative metabolomic biomarkers associated with clinical outcomes (p val < 0.05 ; q val < 0.2) were identified. Of which, eleven were associated with week 14 remission, and the remaining sixteen were associated with non-remission. Ten of the eleven metabolomic features associated with early remission were categorized as either carboxylic acids ($n = 4$), steroid derivatives ($n = 4$) and tetrapyrroles ($n=2$), whereas the majority of metabolomic features associated with non-remission (9/16) were either glycerolipids or glycerophospholipids (Figure 3A).

We performed enrichment analysis, including all 63 differential metabolite features with nominal p -value < 0.1 (Table S3), to profile broad classes of compounds that were significantly overabundant in patients who achieved week 14 remission compared to non-remitters (Figure 3B). Patients who did not achieve week 14 remission, had significant enrichment of serum glycerophosphoethanolamines ($q < 0.01$) and diacylglycerols ($q = 0.03$). Patients with week 14 remission had significant enrichment of serum bile acids (enrichment ratio 1.85, $q < 0.01$), specifically secondary bile acids, such as glycolithocholate (log2-fold difference: 1.66, $p = 0.04$), glycodeoxycholate (log2-fold difference: 1.35, $p = 0.04$), ursodeoxycholate (log2-fold difference: 1.21, $p = 0.03$) in baseline serum samples (Figure 3C).

The gut metagenome may be a source of secondary bile acid production influencing clinical outcomes

The conversion from primary to secondary bile acids is a function restricted to a narrow range of clostridial species and mediated by $7\alpha/\beta$ -dehydroxylation enzymes (Funabashi et al., 2020), of which BaiB, BaiCD, BaiA2, BaiE, BaiF and BaiH were previously shown to be sufficient with no additional enzymes required. Although we have already revealed a few microbial EC features contributing to secondary bile acid biosynthesis that are significantly overabundant amongst patients who achieve week 14 remission on anti-cytokine therapy, we investigated the role of the gut microbiome in determining bile acid metabolite concentrations in greater detail. We mapped the gut metagenome of patients against the recently published *de novo* assembled gene catalogue, (Kenny et al., 2020) which grouped clusters of homologous proteins to uncover homologs of the 8 enzymes making up the *bai* operon, responsible for bile acid $7\alpha/\beta$ -dehydroxylation (Funabashi et al., 2020). Among patients with paired serum and stool samples at baseline, we found significant correlations ($R > 0.5$, p val < 0.01) between the abundance of the *bai* operon enzymes and the relative intensities of secondary bile acids glycolithocholate and glycodeoxycholate (Figure 3D). We subsequently classified individual gut metagenomic samples as either bile acid $7\alpha/\beta$ -dehydroxylation capable (*bai*+) or incapable (*bai*-) (Figure 3E), determined by the presence or absence of set of *bai* operon enzymes described above, and found a larger

fraction of patients with baseline metacommunity DMM group 1, which favored response to anti-cytokine therapy, were classified as *bai* operon positive status (*bai*⁺), compared to patients with baseline DMM group 2 (79.1% vs 32.6%; $p < 0.01$) (Figure 3E). Amongst *bai*⁺ patients (60.2% of the cohort), patients undertaking anti-cytokine therapy were more likely to achieve week 14 remission compared to patients taking anti-integrin therapy (OR 3.52, 95% CI: 1.30–10.1, $p = 0.02$).

Validation of metagenomic and metabolomic associations

To validate the above metagenomic associations with anti-cytokine response, we obtained baseline metagenomic profiles of 46 patients undergoing infliximab therapy from two prior published studies (Lewis et al., 2015; Lloyd-Price et al., 2019). In the validation cohort, 6 of the 9 species associated with anti-cytokine response in our prospective cohort met the abundance threshold (i.e. present in at least 10% of samples with 0.1% relative abundance) for further analysis. Using a 5-fold cross validation random forest classifier on the prospective cohort dataset, the supervised machine learning model performed reasonably well in predicting therapy response in the validation cohort (Accuracy 0.74, 95% CI: 0.59–0.86), whereby majority of these taxonomic features associated with anti-cytokine response in the prospective PRISM cohort ($n = 4/6$; *Eggerthella lenta*, *Collinsella stercoris*, *Gordonibacter pamelaiae*, *Bacteroides ovatus*) were amongst the top 20 features by importance (Figure S4A). We also replicated the association with *bai*⁺ operon positivity and preferential anti-cytokine therapy response. As in the primary cohort, in the independent validation cohort, we found a high proportion of responders to anti-cytokine (76.5%, Figure 3F) amongst patients who were baseline *bai*⁺, and this was associated with a five-fold higher likelihood of clinical remission at week 14, amongst patients on anti-cytokine therapy in the validation cohort, when compared to vedolizumab users in the original manuscript (OR 5.42, 95% CI: 1.59 – 22.1, $p = 0.01$).

Secondary bile acids have been hypothesized to have an anti-inflammatory role in the gut, whereby administration of lithocholic acid was able to ameliorate disease in murine model of colitis (Sinha et al., 2020). To support mechanistic plausibility of the association between presence of the *bai* operon and response to biologic therapy, we examined the association between *bai*⁺ status, fecal secondary bile acid concentrations, and fecal calprotectin measurements using paired fecal metagenomics and metabolomics from 220 participants from previously published studies (Franzosa et al., 2019). We then found that patients with *bai*⁺ status have correspondingly higher abundance of fecal secondary bile acids (deoxycholic acid, tauro-lithocholic acid and tauro-ursodeoxycholic acid; $p < 0.05$, Figure S4B). We further demonstrated that patients with active disease (ie. elevated fecal calprotectin) have lower fecal deoxycholic acid measurements ($p < 0.05$, Figure S4C), whereby fecal calprotectin is an established surrogate marker of gut inflammation (Rubin et al. 2019). These findings support the potential mechanistic role of secondary bile acid concentrations in attenuating luminal inflammation and enhancing biologic therapy response.

Serum cytokines correlated to microbial diversity are associated with remission

To expand our biomarker discovery effort, we profiled 104 unique serum protein biomarkers, which were selected *a priori* based on their direct or indirect role in inflammation and the immune response. These protein markers represent a wide range of gene ontology (GO) immune-related biological functions, such as cytokine-mediated signaling (GO0019221), chemokine-mediated signaling (GO0070098) and cellular response to TNF (GO0033209) (Figure S3B). Using the same multivariate linear model described above for metagenomics and metabolomics, a total of 9 serum proteins were significantly associated with remission outcome in at least one patient group (Figure 4A), after correcting for other covariates (age, subtype of IBD, duration of disease, and medication use). The majority (6/9) of the putative serum proteins were associated with endoscopic remission, rather than clinical remission at either week 14 or at one year. Likewise, a majority of putative proteins were only significantly associated for either patients treated with anti-cytokine (i.e. CD6, CD244, TRAIL, IL18, ADA) or anti-integrin therapy (i.e. NT3, IFNLR1, PIK3AP1). Only CASP8, was found to be significantly associated for both classes of therapeutics, whereby it was positively associated with endoscopic remission for patients taking anti-cytokine therapy ($p_{\text{val}} = 0.02$) and inversely associated with clinical remission with anti-integrin therapy ($p_{\text{val}} < 0.001$, $q_{\text{val}} = 0.16$). Baseline serum TNF, IL-12, IL-22, and IL-23 were not predictive of week 14 or week 52 remission with either therapeutic class.

As the microbiome, in particular metagenomic diversity, was an indicator of anti-cytokine response, we investigated the association of serum immune proteins with prevalent metagenomic taxonomic features, as well as metagenomic indices of alpha diversity. A total of 27 unique proteins were associated with at least one microbial feature at FDR 0.1 (Table S4). The majority of these proteins (25/27) were significantly associated with at least one metagenomic taxonomic feature (Figure 4B, Table S3). Although most of the microbial associations to proteins were previously unknown several have previously been linked to either inflammation or immune response in IBD. Enrichment of *Alistipes* in NOD2 knockout mice, was associated with anti-inflammatory cytokines (TGF- β , IL10) and attenuated 2,4,6-Trinitrobenzene sulfonic acid induced-colitis (Butera et al., 2018). Four proteins were associated with indices of metagenomic alpha diversity (Figure 4C). FGF19 ($R = 0.44$, $q < 0.05$) and ZBTB16 ($R = 0.40$, $q = 0.09$) were positively correlated with microbial richness, whereas IL10RB ($R = -0.42$, $q = 0.08$) and IL12RB1 ($R = -0.39$, $q = 0.08$) were both inversely correlated with Shannon alpha diversity. All 4 of the proteins associated with metagenomic diversity were also differentially expressed between patients with either baseline metacommunities (Figure 4D), whereby patients with baseline metacommunity 1 had increased mean serum expression of FGF19 (log2 fold difference: 0.97, $p = 0.01$) and ZBTB16 (log2 fold difference: 0.65, $p = 0.001$) and patients with baseline metacommunity 2 had increased mean serum expressions of IL10RB (log2 fold difference: 0.20, $p = 0.03$) and IL12RB1 (log2 fold difference: 0.26, $p = 0.03$).

Multi-omics to predict treatment response

Previously we demonstrated amongst patients taking vedolizumab that baseline clinical features alone were insufficient in predicting remission at week 14 (AUC 0.619), and the use of microbial taxa (AUC 0.715) and pathways (AUC 0.738) resulted in improved predictive

ability (Ananthakrishnan et al., 2017). Here, we aim to highlight the incremental value of each layer of information amongst patients taking anti-cytokine therapy using random forest classifiers with feature selection and leave-one-out cross validation; a model using only baseline clinical features performed poorly in predicting week 14 remission with an AUC of 0.624. In contrast, adding metagenomic, metabolomic, or proteomic significantly increased the predictive value with AUCs of 0.849, 0.773, and 0.806, respectively (Figure 4E). Among the 21 participants with available proteomic, metabolomic, and metagenomic profiles at baseline, the AUC of the clinical + metagenomic + metabolomic + proteomic marker in predicting response was 96.3% (95% CI 0.88–1). Furthermore, all three -omic profiles were only available in those initiated on anti-cytokine therapy. Although the integrative analysis was undertaken with internal cross-validation, more robust examinations in larger cohorts are essential prior to application to clinical practice.

DISCUSSION

With availability of multiple therapeutic mechanisms of action, identifying determinants of overall likelihood to improve with treatment, as well as of the more nuanced response to a given treatment class is of high priority in the management of IBD. Here, by applying high throughput -omic profiling to a prospective cohort of patients initiating anti-cytokine or anti-integrin therapy for the management of moderate-to-severe IBD, we identify unique microbial signatures that may preferentially favor response to a specific therapeutic class. We also propose mechanisms for the beneficial effects of a favorable microbial profile.

An important finding from our analysis was the association of microbial metacommunities with distinct likelihoods of response to either anti-integrin or anti-cytokine therapy. The specificity of this association suggests the existence of unique biologic mechanisms that determine response to an individual therapy class, despite broadly similar therapeutic efficacy of antibodies targeting either mechanism. Many of the bacterial species associated with remission after anti-cytokine therapy have well-described anti-inflammatory effects, often by contributing to a favorable fecal short-chain fatty acid profile. Patients more likely to achieve remission with anti-cytokine therapy had a greater abundance of colonic butyrate-producing microbial species which promotes intestinal homeostasis (i.e. *Agathobaculum butyriciproduces*, *Clostridium citroneae*). They also had increased abundance of succinate-consuming microbial species such as *Phascolarctobacterium faecium* (Watanabe et al., 2012). Accumulation of luminal succinate which is primarily microbial in origin can lead to intestinal inflammation. Binding of succinate to the SUCNR1, a G-protein coupled receptor widely expressed across intestinal and immune cells, can act in an autocrine and paracrine manner to increase inflammatory responses in myeloid cells, enhance production of IL-1B and act as a chemotactic factor for dendritic cells, resulting in antigen specific T-cell activation (Connors et al., 2018; Macias-Ceja et al., 2019). Among patients initiating anti-integrin therapy, on the other hand, remitters had increased abundances of *Bifidobacterium longum*, and *Bacteroides species* (*B. ovatus*, *B. stercoris*). Some of the bacterial species that associated with clinical remission have also demonstrated efficacy as standalone microbiome-directed therapy. In a murine colitis model, *B. ovatus* monotherapy was superior to fecal transplantation in achieving clinical remission after DSS administration to C57BL/6 mice (Ihekweazu et al., 2019). In a study of patients with CD, gain of *B.*

longum was associated with maintenance of remission (Kong et al., 2020). Interestingly, the association between the microbiome and clinical response was stronger for week 14 remission than week 52 clinical or endoscopic outcomes. This suggests that the influence of the microbiome may be more dominant on proximate outcomes. A systematic study of different time intervals between microbial profiling and clinical features is important to better identify the optimal predictive window for the microbiome for individual clinical outcomes.

Direct comparison of the metagenomic profiles contributing to our meta-community definitions with other landmark studies using enterotyping, are limited by heterogeneity in the patient population and intent of partitioning into subgroups. Jacobs et al. (2016) used 16S rRNA profiling of populations that contained a mix of patients with IBD and healthy controls to define enterotypes that predicted IBD diagnosis in the study population, or a predisposition towards developing IBD in healthy first-degree relatives. Costea et al. (2018) and Wu et al. (2011) also used 16S rRNA profiling in healthy individuals to define enterotypes and microbial species favoring a specific enterotype. However, neither of the two latter studies included patients with IBD. To our knowledge, there are no prior studies using metagenomic data to partition patients with moderate to severe IBD into distinct microbial communities with implications for clinical phenotype such as therapy response. The metagenomic profile in our cohort reveals both the effect of chronic inflammation on the microbiome in comparison to the microbial features informing enterotype definition described in prior studies (Arumugam et al., 2011), but also species level differences within a genus that highlights the importance of metagenomic sequencing studies.

Serum metabolomic profiling also yielded insight into a patient's potential response to treatment. We observed a strong association between serum secondary bile acid concentrations and likelihood of achieving remission with anti-cytokine therapy. Secondary bile acid induced stimulation of the FXR is protective in chemically induced colitis models (Sinha et al., 2020). We also demonstrated that modification of secondary bile acid profiles may be a mechanism by which the gut microbial composition determines response to biologic therapy. This may occur via a group of Bai 7 α /β-dehydroxylation enzymes. Microbial metacommunities with increased abundance of Bai 7 α /β-dehydroxylation component microbial micro-organisms are more likely to be found in patients responsive to anti-cytokine therapy. This may explain why increased diversity (and therefore more microbes contributing these enzymes) also associates with a positive response. An association between microbial diversity and treatment response is consistent with other observations that diversity is associated with superior rates of remission. Microbial diversity was also inversely associated with serum cytokines profiles (i.e. IL10RB, IL12RB1) that favored resistance rather than response to anti-cytokine therapy.

Similarly, serum proteins associated with remission for either patients treated with anti-cytokine (i.e. CD6, CD244, TRAIL, IL18, ADA, Caspase 8) or anti-integrin therapy (i.e. NT3, IFNLR1, PIK3AP1) tended to be specific to each individual class. The anti-inflammatory roles of some of these cytokines are well established. For example, TRAIL, a transmembrane protein belonging to the TNF superfamily, suppresses gut inflammation and inhibits colitogenic T-cell activation via apoptosis-independent pathways (Chyuan et

al., 2019). The association between Caspase 8 and response to anti-cytokine therapy highlights specific TNF-mediated pathways mediating remission amongst patients on anti-cytokine therapy. Caspase 8 regulates intestinal homeostasis, protecting intestinal epithelial cells from TNF- α induced necroptotic cell death (Günther et al., 2011). In contrast, the association between IFNLR1 and response to anti-integrin therapy suggests a more dominant role for neutrophil migration in mediating inflammation in these patients. IFNLR1, is usually expressed by neutrophils within the mucosal lamina propria to mediate the anti-inflammatory effects of IFN γ . IFNLR1 knock out mice demonstrated more severe intestinal inflammation than wild-type mice when exposed to DSS, exemplifying its anti-inflammatory activity (Broggi et al., 2017). Some of the associations identified, such as IL18, which has been shown here to be positively associated with mucosal healing, are different from observed literature and require the need for wider replication studies.

One limitation of our study was that remission relied primarily on clinical indicators with fewer than half of the cohort having endoscopic outcomes available. Furthermore, not all patients provided both stool and serum at baseline, limiting our statistical power. Despite the study limitations, there are several important implications to our findings. First, we demonstrate that microbial signals that mediate response to immunosuppressive treatment vary by therapeutic class. Harnessing this variability through precision-medicine approaches would allow for tailored therapy at the individual patient level to ensure early response and prevent protracted morbidity due to sequential empiric treatment trials. Our study findings highlight a role not just for microbial taxonomy but for microbially-encoded pathways that may be relevant to treatment. These pathways have been validated by changes in abundance of serum metabolites and corresponding cytokine readouts. Specifically, our study highlights an important role of the microbiome in determining luminal short-chain fatty acid and succinate concentrations that may determine the course of inflammation. Our findings also demonstrate a role for bile acid metabolism, particularly secondary bile acids, as a determinant of response to anti-cytokine therapy in IBD. Identification of predictive microbial signals could also allow for refinement of novel probiotics that may deliver specific anti-inflammatory functions, or stimulate anti-inflammatory metabolic pathways that may act in ameliorating gut inflammation. The 'omics profiles identified in this study could serve as targets for newer therapies and shed further light on the pathogenesis and progression of these complex diseases.

STAR METHODS

RESOURCE AVAILABILITY

Lead Contact: Ramnik J. Xavier (xavier@molbio.mgh.harvard.edu).

Materials Availability: Materials generated in this study are available from the the Lead Contact without restriction.

Data and Code availability: PRISM metagenomics data is available in the Sequence Read Archive (SRA) (<https://www.ncbi.nlm.nih.gov/sra>): PRJNA685168.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Study cohort—This study was nested within a longitudinal prospective IBD cohort at Massachusetts General Hospital (Prospective Registry of IBD study at MGH (PRISM). Details of this cohort have been published previously (Ananthakrishnan et al., 2014). In brief the PRISM registry is open to all adult patients with IBD seeking care at the MGH Crohn's and Colitis center. This nested study population consisted of patients with moderate-to-severe CD or UC initiating anti-TNF (infliximab, adalimumab, golimumab, certolizumab pegol), anti-IL12/23 (ustekinumab), or anti-integrin (vedolizumab) therapy. The former two therapeutic classes were grouped together as anti-cytokine therapy. Characteristics of the included patients initiating either biologic therapy is presented in Supplementary Table 1. Patients initiating therapy were approached for participation in the study prior to therapy initiation. Upon provision of informed consent, detailed information was obtained about disease phenotype according to the Montreal classification, current and prior treatment history including medical and surgical interventions. Concomitant medications at baseline were noted including corticosteroids and antibiotic use. Laboratory values including hemoglobin, serum albumin, C-reactive protein, and erythrocyte sedimentation rate were noted. Patients with a stoma or with an ileo-anal pouch were excluded. Patients were invited to provide blood and stool samples at baseline. Blood samples were centrifuged and separated into buffy coat and serum before storage at -80°C . Stool samples were collected in 100% ethanol, aliquoted, and stored at -80°C prior to DNA extraction and metagenomic sequencing.

Prospective follow-up and definition of outcomes—Enrolled patients completed assessment of disease activity at baseline using validated disease activity scores utilized in clinical trials. These included the Harvey Bradshaw index (HBI) for CD (Harvey and Bradshaw, 1980) and simple clinical colitis activity index (SCCAI) for UC (Walmsley et al., 1998). Upon prospective follow-up, patients completed these scores again at weeks 14, 30, and 54 after therapy initiation. A range of ± 2 weeks was allowed around each time point for completion of these scores. Consistent with prior definitions, HBI or SCCAI scores ≤ 2 indicated clinical remission while scores above this cut-off suggested active disease. At each follow-up visits' concomitant medications were again noted as were whether IBD-related surgeries or hospitalizations had occurred in the interim.

Our primary outcome was achievement of clinical remission at week 14 with a secondary outcome being clinical remission at week 52. For patients undergoing standard of care follow-up colonoscopy to assess mucosal response, we noted whether they had achieved endoscopic remission. This was defined as Simple Endoscopic Score-CD (SES-CD) < 3 without any ulcerations for CD (Daperno et al., 2004) or a Mayo endoscopic score of 0 or 1 for UC (Travis et al., 2013). Fecal calprotectin measurement was not widely available or reimbursed during the study period and was not systematically utilized.

Ethics Statement—This study was approved by the Institutional Review Board of Partners Healthcare and all patients provided informed consent.

METHOD DETAILS

Metagenomic Sequencing—DNA was extracted from ~100 mg of stool using enzymatic treatment with 15 mg/ml lysozyme and proteinase K (Qiagen) for 10 min incubation followed by mechanical lysis using 0.1mm glass beads and 3 min bead beating on the “homogenize” setting of a Mini Bead beater-8 (Biospec Products). After lysis, sample was centrifuged for 5 minutes at maximum speed to pellet debris and supernatant passed through a QIAshredder spin column (Qiagen) prior to proceeding with nucleic acid extraction using an Allprep DNA/RNA Mini Kit (Qiagen) following manufacturer protocol. Metagenomic libraries were constructed from 100–250 pg of DNA using the Nextera XT DNA Library Preparation Kit (Illumina) as per the manufacturer’s recommended protocol. Sequencing was performed on the HiSeq 2500 2x 101 PE platform (Illumina), targeting ~2.5 Gb of sequence per aliquot. Demultiplexing, BAM and FASTQ file generation were performed using the Picard suite (<https://broadinstitute.github.io/picard>).

Metabolomic sequencing—Serum metabolomic profiling was performed at the Broad Institute on the Broad Metabolomics platform (Franzosa et al., 2019) Four liquid chromatography mass-spectrometry (LC-MS) methods were used to profile the following sets of metabolites. Method 1: Amino acids, acylcarnitines, dipeptides, nucleotides, and other cationic polar metabolites measured in 10 µL of serum using HILIC chromatography coupled with untargeted positive ion mode MS. Method 2: Polar and non-polar lipids, including tri-, di-, and mono-acylglycerols; phosphatidyl-cholines (PCs), -ethanolamines (PEs), and -serines; lyso-PCs and lyso-PEs; plasmalogens; cholesterol esters; ceramides; and sphingomyelins measured in 10 µL of serum using C8 chromatography and untargeted positive ion mode MS. Method 3: Short-, medium-, and long-chain free fatty acids, bile acids, and metabolites of intermediate polarity measured in 30 µL of serum using C18 chromatography and untargeted negative ion mode MS. Method 4: Central metabolites, sugars, sugar phosphates, organic acids, purines, and other anionic polar metabolites of known ID measured in 30 µL of serum using HILIC chromatography and untargeted negative ion mode MS. Known metabolites were processed using TraceFinder (v 3.0, Thermo Scientific) software by matching retention times and mass-to-charge ratio (m/z) to synthetic mixtures of reference compounds and characterized pooled plasma reference samples in each sample queue. The untargeted datasets were processed using Progenesis QI software (v 1.0, Nonlinear Dynamics). Data was standardized by relative LC-MS peak areas compared to the nearest study sample in the queue and metabolite profile data was log-transformed prior to analysis. Only 541 unique metabolites features were retained for subsequent analysis, after filtering for at least 50% presence within our cohort, and those matched with internal reference standards that were previously characterized using the Broad Institute methods.

Proteomic sequencing—Serum proteomic profiling was performed at Olink Proteomics (Watertown, MA) using the multiplex PEA technology (Assarsson et al., 2014). Two panels – inflammation and immune response – each comprising 92 biomarkers were run, testing 96 samples simultaneously in each run. The multiplex PEA technology uses paired oligonucleotide-conjugated antibodies as a probe for each protein. When the probes bind to a common target protein, the DNA oligonucleotides hybridize, allowing enzymatic DNA

polymerization to produce a new DNA amplicon molecule which is quantified using a microfluidic real-time polymerase chain reaction (PCR) platform. The data for each protein was expressed as a normalized protein expression (NPX) value, which is an arbitrary unit on a log2 scale allowing for normalization to minimize inter- and intra-assay variation. Only 104 unique proteomic features with at least 40% of readouts above the minimal lower limit of detection were retained for subsequent analysis.

QUANTIFICATION AND STATISTICAL ANALYSIS

Metagenomic profiling—Metagenomic taxonomic and functional profiles were generated using the bioBakery meta’omics workflow (McIver et al., 2018) in the Terra workspace. In brief, reads mapping to the human genome were first filtered out using KneadData (<https://github.com/biobakery/kneaddata>). Taxonomic profiles of shotgun metagenomes were generated using MetaPhlan3 (Beghini et al., 2020), which uses a library of clade-specific markers to provide pan-microbial profiling. Functional profiling was performed by HUMAnN3 (Franzosa et al., 2018), whereby HUMAnN3 constructs a sample-specific reference database from the pangenomes of the subset of species detected in the samples by MetaPhlan3 (pangenomes are precomputed represented of the open reading frames of a given species). Sample reads are mapped against this database to quantify gene presence and abundance on a per-species basis. A translated search is then performed against a UniRef-based protein sequence catalogue (UniRef release 2014_7) for all reads that fail to map at the nucleotide level. The results are abundance profiles of gene families (UniRef90s), for both metagenomics, stratified by each species contributing to those genres, and further summarized to higher level gene groups such as ECs, KOs, or metaCYC pathways. This resulted in the generation of 506 microbial species (from 193 genera) and 2371 annotated EC features for downstream analyses. For subsequent analysis, read counts were transformed into relative abundances by normalization to the total number of reads per sample. Low-abundance filters were applied to discard taxonomic and functional features whose relative abundance did not reach 0.1% and 0.001% respectively, in at least 10% of the individuals.

Analysis of metagenomic diversity and metacommunities—Alpha diversity (richness and Shannon diversity) was calculated for taxonomic profiles at the species level, using the Vegan package in R. Subsequent analysis to determine the cluster of metagenomes (at the genus level) associated with remission, utilized Dirichlet multinomial (DMM) algorithms, whereby DMM bins samples on the basis of microbial community structure, and the appropriate number of clusters was determined based on the lowest Laplace approximation score. We further validated the optimal number of clusters using a partitioning around medoid (PAM)-based approach (see Fig S3F). Differential abundance analyses were performed using the LEfSe algorithm (Segata et al., 2011) to identify significant species taxonomic features between DMM metacommunities with a LDA coefficient threshold of 2.5 and FDR correction threshold of 0.05. For association of metagenomic diversity and metacommunities with categorical and continuous clinical metadata, we applied Fisher’s exact tests, Mann-Whiteley *U*-tests, and logistic regression models when appropriate.

Differential abundance analysis of multi'omic features—We used linear models to identify putative differential abundance analysis of all omic measurement types. Features were log transformed to variance stabilize the data. Zero values were additively smoothed by half of the minimal abundance for each feature. Abundance were then log-transformed and fitted with the following per-feature linear mixed-effects model:

feature ~ (intercept) + age + subtype of IBD + duration of disease + antibiotic use + steroids use + immunomodulator use + prior anti-TNF use + biologic class/remission status

That is, in each per-feature multivariable model, the transformed abundance of each feature was modelled as a function of the remission status nested binary variable (with no remission sued as reference) within each biologic class (anti-cytokine, anti-integrin therapy), while adjusting for age (continuous variable), duration of disease (continuous variables), subtype of IBD (with CD as reference), and concomitant antibiotic, steroids and immunomodulator use (all binary covariates) as well as prior anti-TNF exposure (binary covariate). Fitting was performed with the MaasLin2 package in R, with nominal p-value adjusted for multiple hypothesis testing with a target FDR of 0.2.

Metabolite chemical sub-class enrichment was performed using MetaboAnalyst 4.0 (Chong et al., 2019), with their 1072 subclass metabolite library. Enrichment ratio is computed by the ratio of observed hits over the expected hits.

Metagenomic profiling of Bai enzymes—To detect the presence of microbial enzymes of the *bai* operon, metagenomic sequences were mapped to the previously published gene catalogue that was assembled from geographically diverse 3097 gut metagenomes, including PRIM and HMP2 cohorts (Kenny et al., 2020). In brief, this gene catalogue consisted of metagenomic reads from all above cohorts assembled into contigs using MegaHIT, followed by open reading frame prediction with Prodigal and a non-redundant gene catalogue was constructed by clustering predicted genes based on sequence similarity of 95% and 90% coverage of the shorter sequences using CD-HIT. Reads after mapping to the gene catalogue with BWA were filtered by at least 95% sequence identity over the read length, and normalized to transcript-per-million (TPM). USEARCH ublast was used to identify in the gene catalogue sequences homologous to genes in the bile acid inducible operon encoded by *Clostridium scindens* (Funabashi et al., 2020) at the min. 40% AA identity and 80% sequence coverage.

Correlation between multi'omic features—To study the relationship between metagenomic Bai enzymes and serum secondary bile acids profiles, we used spearman correlation to associate the abundance of each Bai enzyme homolog TPM (log10 transform) with the log2 transformed serum relative intensities of glycolithocholate acid and glycodeoxycholate acid. *Bai* presence status (i.e. *bai*+) for each metagenomic sample, was then coded as a dichotomous variable based on the presence (TPM >0) of all of the following: BaiB, BaiCD, BaiA2, BaiE, BaiF and BaiH (which was previously shown to be sufficient with no additional enzymes required for 7 α -hydroxylation). This dichotomous variable (Bai+/Bai-) was then used for subsequent comparison of distribution between DMM metacommunities, as well as, odds for remission between patients treated with either

anti-cytokine therapy or anti-integrin therapy, using fisher exact and logistic regression respectively.

To study the relationship between metagenomic features and serum proteins, we again used the spearman correlation to associate the log-transformed metagenomic taxonomic features with NPX-normalized serum cytokines, and corrected for multiple hypothesis testing using a FDR cut off of 0.1.

Validation cohorts—To validate the results from our prospective cohort, we used published data from two separate studies. First, to serve as an independent validation cohort for the metagenomic associations with therapy response, we used published clinical and metagenomic data from Lewis et al. (2015) that examined outcomes after infliximab initiation in 33 children with Crohn’s disease and data from Lloyd Price et al. (2019) that included 13 patients with moderate-to-severe IBD initiating infliximab therapy on whom disease activity assessment using the SCCAI or HBI at 14 weeks after therapy initiation was available. Almost all of the validation cohort had Crohn’s disease (n = 45/46), with only one having ulcerative colitis. The week 14 remission rate was 65.2%, slightly higher than our prospective cohort. Raw sequences from the validation dataset were parsed through the same metagenomic pipeline outlined above, using both the reference based Metaphlan3 and the previously published *de novo* assembled gene catalogue to map for Bai enzyme homologs.

To confirm the association between secondary bile acid concentrations and reduced inflammation (higher rate of response to therapy), we used published data from two additional independent cohorts of patients with IBD, previously described by Franzosa et al (2019). This included 155 patients from the MGH PRISM cohort and 65 patients from two separate cohorts in the Netherlands. Similarly, annotated fecal metabolomic abundances, fecal calprotectin data and raw sequences were obtained. Raw sequences were again parsed through the same *de novo* assembled gene catalogue previously described to map for Bai enzyme homologs. Fecal measurements of secondary bile acids (deoxycholic acid, tauro-lithocholic acid, ursodeoxycholic acid) were then compared between *bai*⁺ and *bai*[−] patients. Subsequently secondary bile acids, in particular deoxycholic acid, was compared between active and inactive IBD patients, using paired fecal calprotectin measurements, as a surrogate marker of luminal inflammation with varying clinical cut-offs (i.e. <50, <100, <150) (Rubin et al., 2019).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

The authors thank participating patients and research staff at the MGH Crohn’s and Colitis center. The authors also thank Luke Besse for project management and making data available through the Sequence Read Archive (SRA) and both the Broad Institute Microbial ‘Omics Core and Broad Metabolomics Core for help with sequencing data generation. The authors are also grateful to Heather Kang for editorial assistance with the manuscript and figures, as well as Melanie Schirmer for helpful discussion. This study was funded by the National Institutes of Health (RO1 AT009708, P30 DK043351, U19 AI142784, U19 AI110495 to RJX; R21 DK127227 to ANA), Crohn’s and Colitis Foundation Microbiome Initiative (RJX), Crohn’s and Colitis Foundation Senior Research Award (ANA), Center

for Microbiome Informatics and Therapeutics (RJK), Chleck Family Foundation (ANA), and the National Medical Research Council (JWJL).

REFERENCES

- Aden K, Rehman A, Waschina S, Pan W-H, Walker A, Lucio M, Nunez AM, Bharti R, Zimmerman J, Bethge J, et al. (2019). Metabolic Functions of Gut Microbes Associate With Efficacy of Tumor Necrosis Factor Antagonists in Patients With Inflammatory Bowel Diseases. *Gastroenterology* 157, 1279–1292.e11. [PubMed: 31326413]
- Ananthakrishnan AN, Huang H, Nguyen DD, Sauk J, Yajnik V, and Xavier RJ (2014). Differential effect of genetic burden on disease phenotypes in Crohn's disease and ulcerative colitis: analysis of a North American cohort. *Am. J. Gastroenterol* 109, 395–400. [PubMed: 24419484]
- Ananthakrishnan AN, Luo C, Yajnik V, Khalili H, Garber JJ, Stevens BW, Cleland T, and Xavier RJ (2017). Gut Microbiome Function Predicts Response to Anti-integrin Biologic Therapy in Inflammatory Bowel Diseases. *Cell Host Microbe* 21, 603–610.e3. [PubMed: 28494241]
- Ananthakrishnan AN, Kaplan GG, and Ng SC (2020). Changing Global Epidemiology of Inflammatory Bowel Diseases: Sustaining Health Care Delivery Into the 21st Century. *Clin. Gastroenterol. Hepatol* 18, 1252–1260. [PubMed: 32007542]
- Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, Fernandes GR, Tap J, Bruls T, Batto J-M, et al. (2011). Enterotypes of the human gut microbiome. *Nature* 473, 174–180. [PubMed: 21508958]
- Assarsson E, Lundberg M, Holmquist G, Björkstén J, Thorsen SB, Ekman D, Eriksson A, Rennel Dickens E, Ohlsson S, Edfeldt G, et al. (2014). Homogenous 96-plex PEA immunoassay exhibiting high sensitivity, specificity, and excellent scalability. *PLoS One* 9, e95192. [PubMed: 24755770]
- Beghini F, McIver LJ, Blanco-Míguez A, Dubois L, Asnicar F, Maharjan S, Mailyan A, Thomas AM, Manghi P, Valles-Colomer M, et al. (2021). Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. *Elife* 10.
- Bolger AM, Lohse M, and Usadel B (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. [PubMed: 24695404]
- Broggi A, Tan Y, Granucci F, and Zanoni I (2017). IFN- λ suppresses intestinal inflammation by non-translational regulation of neutrophil function. *Nat. Immunol* 18, 1084–1093. [PubMed: 28846084]
- Butera A, Di Paola M, Pavarini L, Strati F, Pindo M, Sanchez M, Cavalieri D, Boirivant M, and De Filippo C (2018). Nod2 Deficiency in mice is Associated with Microbiota Variation Favouring the Expansion of mucosal CD4⁺ LAP⁺ Regulatory Cells. *Sci. Rep* 8, 14241. [PubMed: 30250234]
- Chong J, Wishart DS, and Xia J (2019). Using MetaboAnalyst 4.0 for Comprehensive and Integrative Metabolomics Data Analysis. *Curr. Protoc. Bioinformatics* 68, e86. [PubMed: 31756036]
- Chyuan IT, Tsai HF, Wu CS, and Hsu PN (2019). TRAIL suppresses gut inflammation and inhibits colitogenic T-cell activation in experimental colitis via an apoptosis-independent pathway. *Mucosal Immunol* 12, 980–989. [PubMed: 31076664]
- Connors J, Dawe N, and Van Limbergen J (2018). The Role of Succinate in the Regulation of Intestinal Inflammation. *Nutrients* 11.
- Costea PI, Hildebrand F, Arumugam M, Bäckhed F, Blaser MJ, Bushman FD, de Vos WM, Ehrlich SD, Fraser CM, Hattori M, et al. (2018). Enterotypes in the landscape of gut microbial community composition. *Nat Microbiol* 3, 8–16. [PubMed: 29255284]
- Danese S, Vuitton L, and Peyrin-Biroulet L (2015). Biologic agents for IBD: practical insights. *Nat. Rev. Gastroenterol. Hepatol* 12, 537–545. [PubMed: 26284562]
- Daperno M, D'Haens G, Van Assche G, Baert F, Bulois P, Maunoury V, Sostegni R, Rocca R, Pera A, Gevers A, et al. (2004). Development and validation of a new, simplified endoscopic activity score for Crohn's disease: the SES-CD. *Gastrointest. Endosc* 60, 505–512.
- Edgar RC (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460–2461. [PubMed: 20709691]
- Estevinho MM, Rocha C, Correia L, Lago P, Ministro P, Portela F, Trindade E, Afonso J, Peyrin-Biroulet L, and Magro F (2020). Features of Fecal and Colon Microbiomes Associate With

- Responses to Biologic Therapies for Inflammatory Bowel Diseases: A Systematic Review. *Clin. Gastroenterol. Hepatol* 18, 1054–1069. [PubMed: 31526845]
- Franzosa EA, McIver LJ, Rahnavard G, Thompson LR, Schirmer M, Weingart G, Lipson KS, Knight R, Caporaso JG, Segata N, et al. (2018). Species-level functional profiling of metagenomes and metatranscriptomes. *Nat. Methods* 15, 962–968. [PubMed: 30377376]
- Franzosa EA, Sirota-Madi A, Avila-Pacheco J, Fornelos N, Haiser HJ, Reinker S, Vatanen T, Hall AB, Mallick H, McIver LJ, et al. (2019). Gut microbiome structure and metabolic activity in inflammatory bowel disease. *Nat Microbiol* 4, 293–305. [PubMed: 30531976]
- Fu J, Wei B, Wen T, Johansson MEV, Liu X, Bradford E, Thomsson KA, McGee S, Mansour L, Tong M, et al. (2011). Loss of intestinal core 1-derived O-glycans causes spontaneous colitis in mice. *J. Clin. Invest* 121, 1657–1666. [PubMed: 21383503]
- Funabashi M, Grove TL, Wang M, Varma Y, McFadden ME, Brown LC, Guo C, Higginbottom S, Almo SC, and Fischbach MA (2020). A metabolic pathway for bile acid dehydroxylation by the gut microbiome. *Nature* 582, 566–570. [PubMed: 32555455]
- Gevers D, Kugathasan S, Denson LA, Vázquez-Baeza Y, Van Treuren W, Ren B, Schwager E, Knights D, Song SJ, Yassour M, et al. (2014). The treatment-naïve microbiome in new-onset Crohn's disease. *Cell Host Microbe* 15, 382–392. [PubMed: 24629344]
- Graham DB, and Xavier RJ (2020). Pathway paradigms revealed from the genetics of inflammatory bowel disease. *Nature* 578, 527–539. [PubMed: 32103191]
- Günther C, Martini E, Wittkopf N, Amann K, Weigmann B, Neumann H, Waldner MJ, Hedrick SM, Tenzer S, Neurath MF, et al. (2011). Caspase-8 regulates TNF- α -induced epithelial necroptosis and terminal ileitis. *Nature* 477, 335–339. [PubMed: 21921917]
- Hall AB, Yassour M, Sauk J, Garner A, Jiang X, Arthur T, Lagoudas GK, Vatanen T, Fornelos N, Wilson R, et al. (2017). A novel *Ruminococcus gnavus* clade enriched in inflammatory bowel disease patients. *Genome Med* 9, 103. [PubMed: 29183332]
- Harvey RF, and Bradshaw JM (1980). A simple index of Crohn's-disease activity. *Lancet* 1, 514. [PubMed: 6102236]
- Ihekweazu FD, Fofanova TY, Queliza K, Nagy-Szakal D, Stewart CJ, Engevik MA, Hultén KG, Tatevian N, Graham DY, Versalovic J, et al. (2019). *Bacteroides ovatus* ATCC 8483 monotherapy is superior to traditional fecal transplant and multi-strain bacteriotherapy in a murine colitis model. *Gut Microbes* 10, 504–520. [PubMed: 30663928]
- Jacobs JP, Goudarzi M, Singh N, Tong M, McHardy IH, Ruegger P, Asadourian M, Moon B-H, Ayson A, Borneman J, et al. (2016). A Disease-Associated Microbial and Metabolomics State in Relatives of Pediatric Inflammatory Bowel Disease Patients. *Cell Mol Gastroenterol Hepatol* 2, 750–766. [PubMed: 28174747]
- Kenny DJ, Plichta DR, Shungin D, Koppel N, Hall AB, Fu B, Vasan RS, Shaw SY, Vlamakis H, Balskus EP, et al. (2020). Cholesterol Metabolism by Uncultured Human Gut Bacteria Influences Host Cholesterol Level. *Cell Host Microbe* 28, 245–257.e6. [PubMed: 32544460]
- Kong L, Lloyd-Price J, Vatanen T, Seksik P, Beaugerie L, Simon T, Vlamakis H, Sokol H, and Xavier RJ (2020). Linking strain engraftment in fecal microbiota transplantation with maintenance of remission in Crohn's disease. *Gastroenterology*
- Lewis JD, Chen EZ, Baldassano RN, Otley AR, Griffiths AM, Lee D, Bittinger K, Bailey A, Friedman ES, Hoffmann C, et al. (2015). Inflammation, Antibiotics, and Diet as Environmental Stressors of the Gut Microbiome in Pediatric Crohn's Disease. *Cell Host Microbe* 18, 489–500. [PubMed: 26468751]
- Lloyd-Price J, Arze C, Ananthakrishnan AN, Schirmer M, Avila-Pacheco J, Poon TW, Andrews E, Ajami NJ, Bonham KS, Brislawn CJ, et al. (2019). Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature* 569, 655–662. [PubMed: 31142855]
- Macías-Ceja DC, Ortiz-Masiá D, Salvador P, Gisbert-Ferrándiz L, Hernández C, Hausmann M, Rogler G, Esplugues JV, Hinojosa J, Alós R, et al. (2019). Succinate receptor mediates intestinal inflammation and fibrosis. *Mucosal Immunol* 12, 178–187. [PubMed: 30279517]
- McIver LJ, Abu-Ali G, Franzosa EA, Schwager R, Morgan XC, Waldron L, Segata N, and Huttenhower C (2018). bioBakery: a meta-omic analysis environment. *Bioinformatics* 34, 1235–1237. [PubMed: 29194469]

- Peyrin-Biroulet L, Panés J, Sandborn WJ, Vermeire S, Danese S, Feagan BG, Colombel J-F, Hanauer SB, and Rycroft B (2016). Defining Disease Severity in Inflammatory Bowel Diseases: Current and Future Directions. *Clin. Gastroenterol. Hepatol* 14, 348–354.e17. [PubMed: 26071941]
- Plaza Oñate F, Le Chatelier E, Almeida M, Cervino ACL, Gauthier F, Magoulès F, Ehrlich SD, and Pichaud M (2019). MSPminer: abundance-based reconstitution of microbial pan-genomes from shotgun metagenomic data. *Bioinformatics* 35, 1544–1552. [PubMed: 30252023]
- Plichta DR, Graham DB, Subramanian S, and Xavier RJ (2019). Therapeutic Opportunities in Inflammatory Bowel Disease: Mechanistic Dissection of Host-Microbiome Relationships. *Cell* 178, 1041–1056. [PubMed: 31442399]
- Rhodes JM (2007). The role of *Escherichia coli* in inflammatory bowel disease. *Gut* 56, 610–612. [PubMed: 17440180]
- Rutgeerts P, Goobes K, Peeters M, Hiele M, Penninckx F, Aerts R, Kerremans R, and Vantrappen G (1991). Effect of faecal stream diversion on recurrence of Crohn's disease in the neoterminal ileum. *Lancet* 338, 771–774. [PubMed: 1681159]
- Schirmer M, Denson L, Vlamakis H, Franzosa EA, Thomas S, Gotman NM, Rufo P, Baker SS, Sauer C, Markowitz J, et al. (2018). Compositional and Temporal Changes in the Gut Microbiome of Pediatric Ulcerative Colitis Patients Are Linked to Disease Course. *Cell Host Microbe* 24, 600–610.e4. [PubMed: 30308161]
- Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, and Huttenhower C (2011). Metagenomic biomarker discovery and explanation. *Genome Biol* 12, R60. [PubMed: 21702898]
- Sinha SR, Haileselassie Y, Nguyen LP, Tropini C, Wang M, Becker LS, Sim D, Jarr K, Spear ET, Singh G, et al. (2020). Dysbiosis-Induced Secondary Bile Acid Deficiency Promotes Intestinal Inflammation. *Cell Host Microbe* 27, 659–670.e5. [PubMed: 32101703]
- Travis SPL, Schnell D, Krzeski P, Abreu MT, Altman DG, Colombel J-F, Feagan BG, Hanauer SB, Lichtenstein GR, Marteau PR, et al. (2013). Reliability and initial validation of the ulcerative colitis endoscopic index of severity. *Gastroenterology* 145, 987–995. [PubMed: 23891974]
- Walmsley RS, Ayres RC, Pounder RE, and Allan RN (1998). A simple clinical colitis activity index. *Gut* 43, 29–32. [PubMed: 9771402]
- Watanabe Y, Nagai F, and Morotomi M (2012). Characterization of *Phascolarctobacterium succinatutens* sp. nov., an Asaccharolytic, Succinate-Utilizing Bacterium Isolated from Human Feces. *Appl. Environ. Microbiol* 78, 511. [PubMed: 22081579]
- Wu GD, Chen J, Hoffmann C, Bittinger K, Chen Y-Y, Keilbaugh SA, Bewtra M, Knights D, Walters WA, Knight R, et al. (2011). Linking long-term dietary patterns with gut microbial enterotypes. *Science* 334, 105–108. [PubMed: 21885731]

HIGHLIGHTS

- Multiomics define microbiome mechanisms linked to remission with various IBD therapies.
- 7 α / β -dehydroxylation of primary to secondary bile acids determines anti-cytokine responses
- Metagenomic and immune profiles linked with remission are unique to therapeutic class.
- 27 unique serum immune-proteins are associated with gut microbiome.

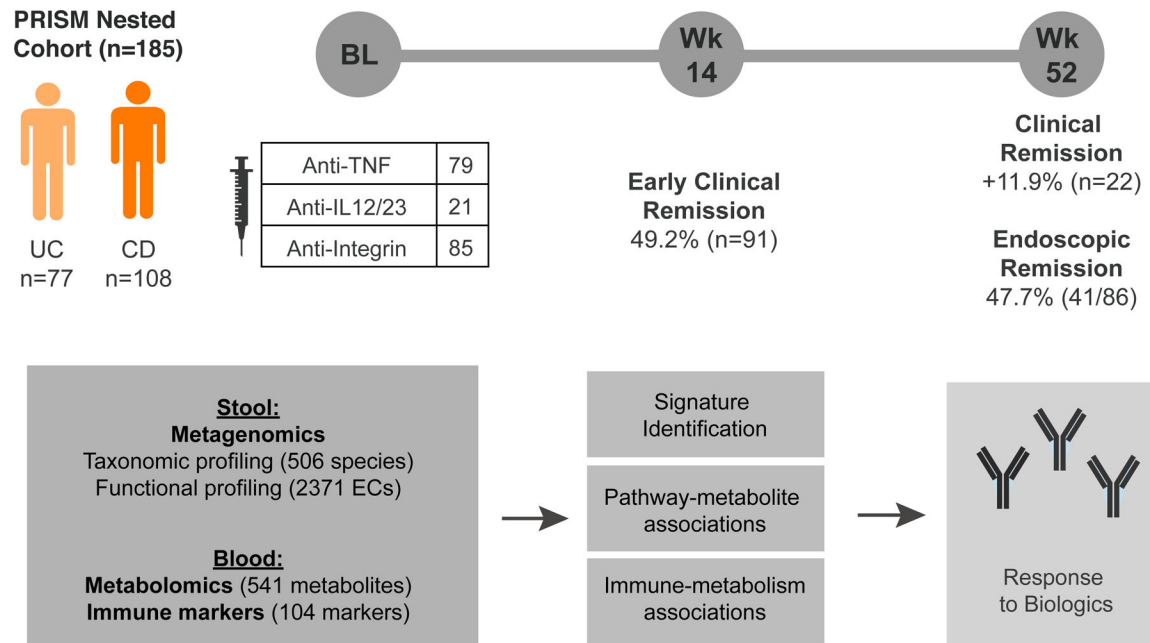


Figure 1. Study overview.

We collected multiple omics measures from IBD patients upon the initiation of one of three different therapeutic classes: anti-TNF, anti-IL12/IL23 or anti-integrin. Outcomes were evaluated at week 14 and week 52. In total, 114 stool samples and 130 serum samples were collected. Each microbial, metabolomic, and proteomic feature was associated with clinical outcomes using feature-wise linear modelling, while adjusting for significant clinical covariates such as age, subtype of IBD, duration of disease, and concomitant medications. Marginal correlation analysis was used to identify significant associations between various ‘omic measures to better define the potential functional mechanisms of the impact of the microbiome on IBD remission.

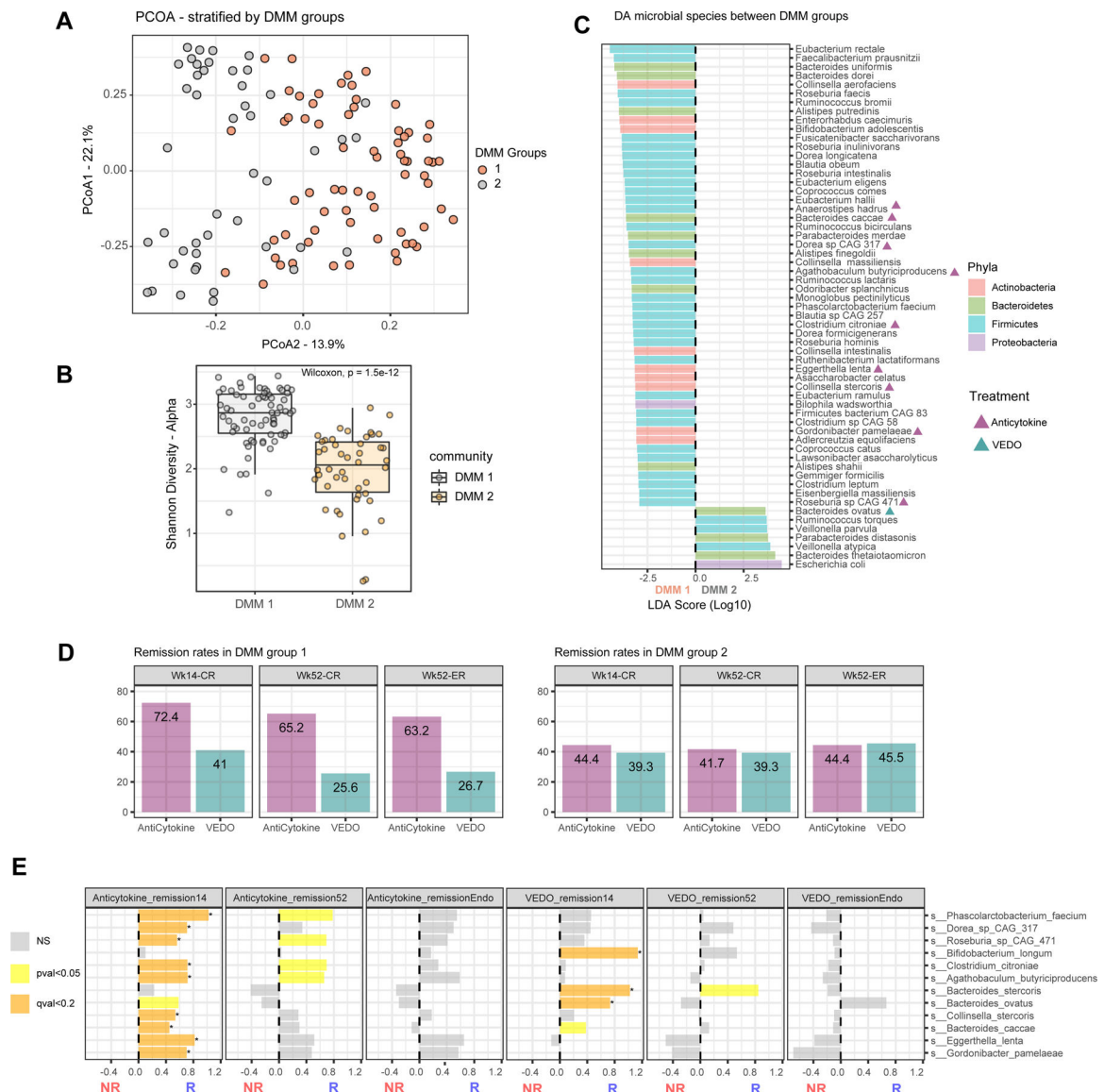


Figure 2. Fecal metagenomics predicts differential response to anti-cytokine and anti-integrin therapy.

(A) Principal coordinate analysis (PCoA) of PRISM cohort subjects based on fecal metagenomic profiles colored by Dirichlet multinomial mixture (DMM) metacommunities.

(B) Boxplot of alpha diversity (Shannon index) plotted stratified by DMM metacommunities. (C) Fitting microbial taxonomic features to DMM models defined two metacommunities; rows represent selected microbial species (arcsine square root-transformed) ranked in descending order of linear discriminant analysis scores from top to bottom.

(D) Differential remission response between those on anti-cytokine therapy and anti-integrin therapy stratified by DMM metacommunities. (E) We applied a general linear model to identify associations ($FDR < 0.2$) of microbial abundance while on anti-cytokine therapy or anti-integrin therapy, with either clinical remission at week 14 or 52, or endoscopic remission. Results are displayed for three different drug subsets and at multiple taxonomic resolutions.

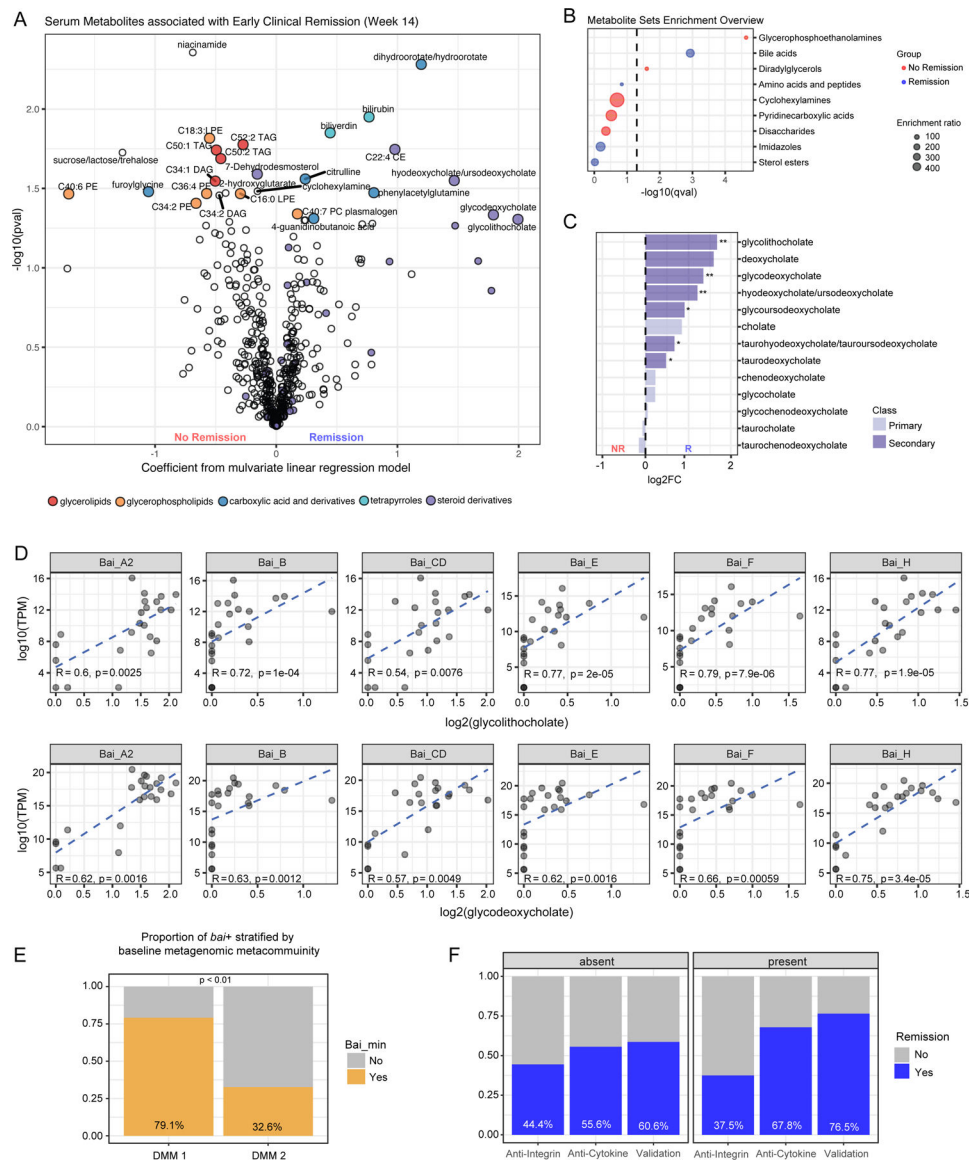


Figure 3. Serum secondary bile acids predict early remission with anti-cytokine therapy. (A) Volcano plot summarizing the effect of baseline serum metabolites associated with week 14 clinical remission. Hits with nominal p-value < 0.05 are labeled, and significant metabolites classified as either glycerolipids (red), glycerophospholipids (orange), carboxylic acid and derivatives (blue), tetrapyrroles (teal) or steroids and steroid derivatives (purple) are colored. (B) Bubble plot summarizing the significant enriched subclasses of metabolites amongst non-remitters (red) or remitters (blue) using MetaboAnalyst4.0, with a vertical dashed line representing qval of 0.05. (C) Bar plot illustrating median log₂ fold change relative intensity of bile acids between non-remitters and remitters. NR, no remission; R, remission. (D) Scatter plot demonstrating strong correlation (Spearman) between abundance of homologous proteins queried with 7 α / β -dehydroxylation Bai enzymes in fecal samples against paired relative intensities of serum secondary bile acids (glycodexoycholate, glycolithocholate). (E) Bar plot demonstrating

the differential distribution of 7 α / β -dehydroxylation Bai competent (*bai*+) metagenomic samples stratified by DMM metacommunities ($p < 0.01$). **(F)** Barplot demonstrating the differential response (blue) stratified by biologic subtype in MGH and external validation cohorts and baseline metagenomic *bai*+ presence.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

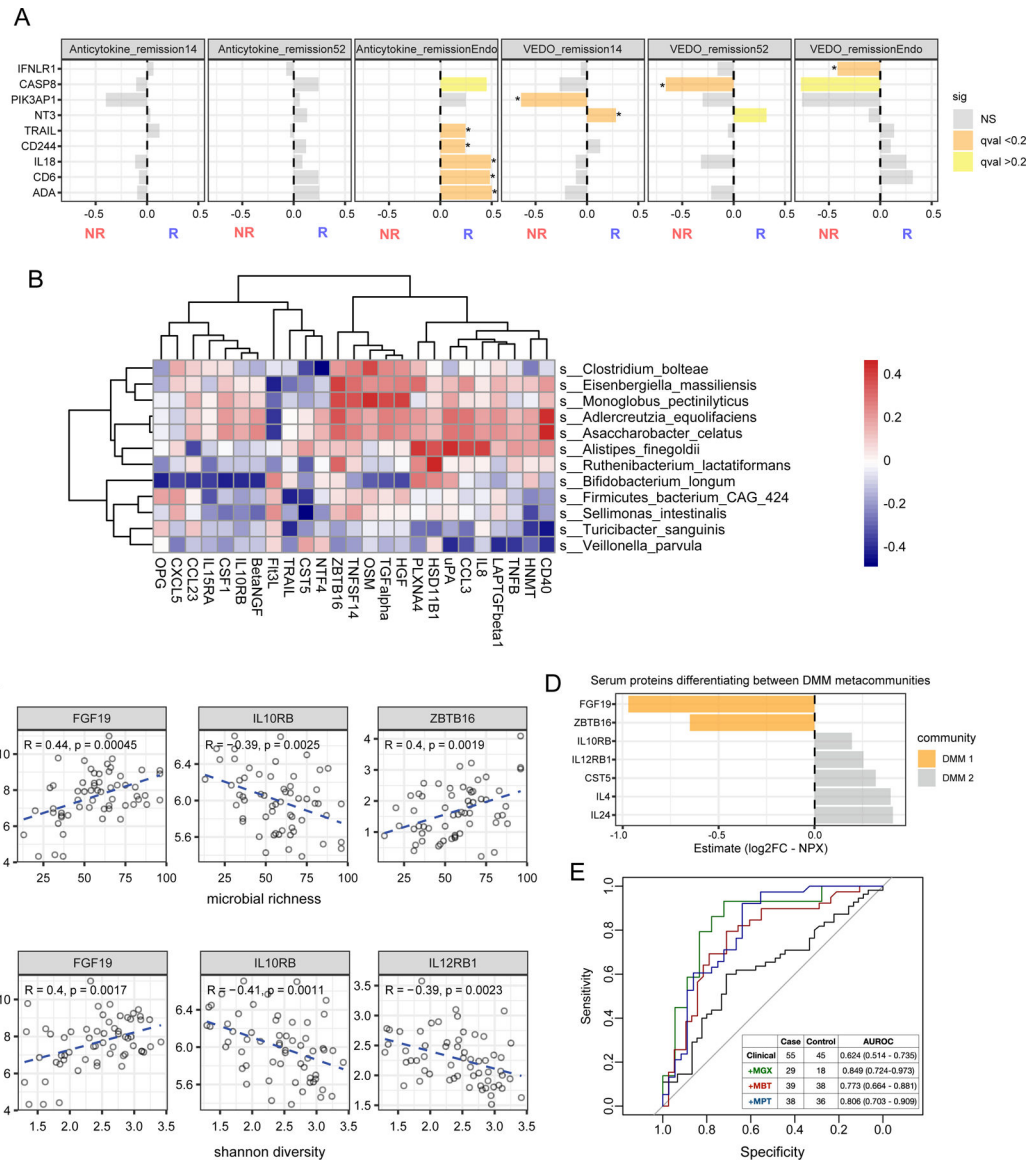


Figure 4. Serum protein markers associated with microbial diversity and clinical remission. (A) Nine proteins associated with a remission outcome in at least one patient group, after multivariate linear regression analysis to identify associations ($FDR < 0.2$) of 104 serum proteomic features while on anti-cytokine therapy or anti-integrin therapy, with either week 14, 52 clinical remission, or endoscopic remission. NR, no remission; R, remission. (B) Heatmap of 25 serum proteins in IBD patients with at least one significant ($q < 0.1$) association with metagenomic taxonomic feature. (C) Four unique serum proteins (FGF19, ZBTB16, IL10RB, IL12RB), expressed as normalized protein expression (NPX) units (see methods), that independently correlate with microbial diversity. (D) Differentially abundant serum proteins associated with baseline metacomunities 1 and 2. (E) The RF classifiers were deployed using a leave-one-out cross validation scheme, whereby the performance of each model in classifying remission at week 14 amongst patients taking anti-cytokine therapy are shown in the AUROC plot, colored with unique colors (black - clinical variables

only; green - clinical and metagenomic features; brown - clinical and metabolomic features; blue - clinical and proteomics, orange - clinical and all three - omics).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

KEY RESOURCE TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological Samples		
IBD patients fecal and serum samples	Crohn's and Colitis Center, Massachusetts General Hospital and Harvard Medical School	PRISM
Critical Commercial Assays		
Mini Bead beater-8	Biospec Products	Cat# 693
QIAshredder	Qiagen	Cat# 79654
Proteinase K	Qiagen	Cat# 19131
AllPrep DNA/RNA Mini Kit	Qiagen	Cat# 80204
Nextera XT DNA Library Preparation kit	Illumina	Cat# FC-131-1096
Deposited Data		
Shotgun metagenomic sequences	Sequence Read Archive	PRJNA685168
Software and Algorithms		
bioBakery meta'omics workflow	(McIver et al., 2018)	https://app.terra.bio/#workspaces/rjxmicrobiome/mtx_workflow
Trimmomatic (v0.36)	(Bolger et al., 2014)	http://www.usadellab.org/cms/?page=trimmomatic
KneadData (v0.7.2)	Huttenhower lab	http://huttenhower.sph.harvard.edu/kneaddata
Metaphlan 3	(Beghini et al., 2020)	https://huttenhower.sph.harvard.edu/metaphlan
HUMAnN3	(Franzosa et al., 2018)	https://huttenhower.sph.harvard.edu/humann
MaAsLin2	Huttenhower lab	https://huttenhower.sph.harvard.edu/maaslin/
MSPminer	(Plaza Oñate et al., 2019)	https://www.enterome.com/downloads/
USEARCH (v8.1)	(Edgar, 2010)	https://www.drive5.com/usearch/
Metabolanalyst 4.0	(Chong et al., 2019)	https://www.metaboanalyst.ca/
R (v3.6.1)	packages for analytical part: vegan, finalfit, tidyverse, broom, qvalue, HMP	https://www.r-project.org/