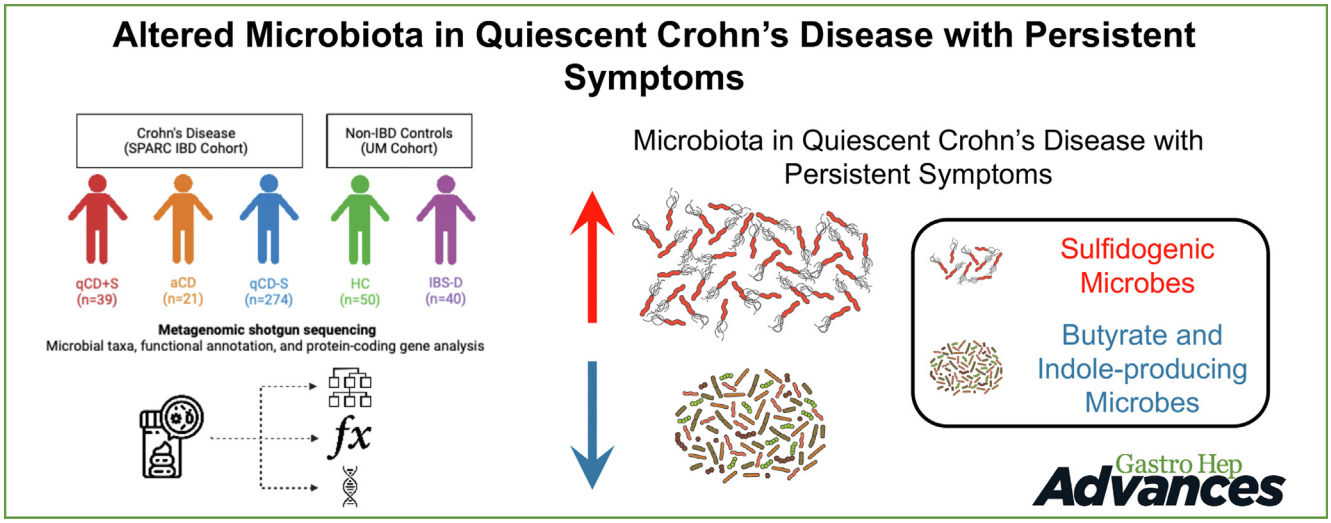
Gastro Hep Advances 2024;3:167–177

ORIGINAL RESEARCH—CLINICAL

# The Microbiome in Quiescent Crohn's Disease With Persistent Symptoms Show Disruptions in Microbial Sulfur and Tryptophan Pathways

Jonathan Golob,[1,*] Krishna Rao,[1,*] Jeffrey A. Berinstein,[2] William D. Chey,[2] Chung Owyang,[2] Nobuhiko Kamada,[2] Peter D. R. Higgins,[2] Vincent Young,[1,3] Shrinivas Bishu,[2,†] and Allen A. Lee[2,†]

[1]*Division of Infectious Diseases, Department of Internal Medicine, University of Michigan, Ann Arbor, Michigan;* [2]*Division of Gastroenterology, Department of Internal Medicine, University of Michigan, Ann Arbor, Michigan; and* [3]*Department of Microbiology and Immunology, University of Michigan, Ann Arbor, Michigan*




**BACKGROUND AND AIMS:** Even in the absence of inflammation, persistent symptoms in Crohn's disease (CD) are prevalent and worsen quality of life. Amongst patients without inflammation (quiescent CD), we hypothesized that microbial community structure and function, including tryptophan metabolism, would differ between patients with persistent symptoms (**qCD + S**) and without persistent symptoms (**qCD-S**). **METHODS:** We performed a multicenter observational study nested within the Study of a Prospective Adult Research Cohort with Inflammatory Bowel Disease. Quiescent inflammation was defined by fecal calprotectin level <150 mcg/g. Persistent symptoms were defined by Crohn's Disease Patient-Reported Outcome-2. Active CD, diarrhea-predominant irritable bowel syndrome, and healthy controls were included as controls. Stool samples underwent whole-genome shotgun metagenomic sequencing. **RESULTS:** Thirty-nine patients with qCD + S, 274 qCD-S, 21 active CD, 40 diarrhea-predominant irritable bowel syndrome, and 50 healthy controls were included for analysis. Patients with qCD + S had a less-diverse microbiome. Furthermore, patients with qCD + S showed significant enrichment of bacterial species that are normal inhabitants of the oral microbiome (eg *Rothia dentocariosa, Fusobacterium nucleatum*) and sulfidogenic microbes (eg *Prevotella copri, Bilophila* spp.). Depletion of important butyrate and indole producers (eg *Eubacterium rectale, Faecalibacterium prausnitzii*) was also noted in qCD + S. Potential metagenome-related functional changes in cysteine and methionine metabolism, ATP transport, and redox reactions were disturbed in qCD + S, also suggestive of altered sulfur metabolism. Finally, qCD + S showed significant reductions in bacterial *tnaA* genes, which mediate tryptophan metabolism to indole, and significant *tnaA* allelic variation compared with qCD-S. **CONCLUSION:** The microbiome in qCD + S showed significant differences in

*Shared co-first authorship. †Shared co-senior authorship.

***Abbreviations used in this paper:*** **aCD, active inflammatory CD; CD, Crohn's disease; CD-PRO2, Crohn's Disease Patient-Reported Outcome-2; FCP, fecal calprotectin; HCs, healthy controls; IBD, inflammatory bowel disease; IBS-D, diarrhea-predominant irritable bowel syndrome; KO, KEGG Orthology; PRO, patient-reported outcome; qCD, quiescent Crohn's disease; qCD + S, qCD with persistent symptoms; qCD-S, qCD without persistent symptoms; SPARC IBD, Study of a Prospective Adult Research Cohort with Inflammatory Bowel Disease.**

Most current article



2772-5723
https://doi.org/10.1016/j.gastha.2023.11.005

sulfidogenesis, butyrate producers, and typically oral microbes compared to qCD-S and active CD. These results suggest that inflammation may lead to durable microbiome alterations which may mediate persistent symptoms through testable mechanisms.

*Keywords:* Quiescent Crohn's Disease; Hydrogen Sulfide; Tryptophan; Patient-Reported Outcomes; Gut Microbiome

## Introduction

Although modern therapies have improved care in inflammatory bowel disease (IBD), improving inflammation does not necessarily improve quality of life in many IBD patients. Even in the absence of active inflammation (ie, "quiescent"), persistent gastrointestinal symptoms are reported in up to 46% of IBD patients, particularly those with Crohn's disease (CD).[1] Quiescent CD with persistent gastrointestinal symptoms (**qCD + S**) result in significant reduction in quality of life similar to active inflammation.[2] Patients with qCD + S are at higher risk for opioid use,[3] exhibit higher mortality,[4] and have higher costs of IBD care.[5] Furthermore, the Food and Drug Administration and the European Medicines Agency recommend use of patient-reported outcome (PRO) measures as a primary end point in IBD clinical trials, which highlights the importance of understanding drivers of qCD + S. Despite the importance of this issue, the mechanisms underpinning this condition are poorly understood, which limit our therapeutic approaches.

Prior studies have demonstrated disrupted epithelial barrier integrity in qCD + S.[6,7] Although initially attributed to subclinical inflammation, the prevalence of persistent symptoms is similar in quiescent IBD patients with and without deep remission.[8] Thus, inflammation alone cannot fully explain qCD + S. More recently, altered microbial communities have been described in quiescent IBD with persistent diarrhea,[9] psychological distress,[10] and fatigue.[11] However, the mechanisms by which an altered microbiome may mediate qCD + S are still largely unexplored. One potential mechanism was suggested in a prior study where quiescent IBD patients with fatigue had decreased serum tryptophan levels.[11] As tryptophan is important in regulating host-microbial interactions and intestinal barrier function, tryptophan may be an important mediator of qCD + S.[12] Thus, we hypothesized that qCD + S would exhibit altered microbial metabolism, particularly with microbial genes important in tryptophan metabolism, compared to quiescent CD without symptoms (**qCD-S**).

## Materials and Methods

### Patient Cohort

We performed a multicenter observational study nested within the Study of a Prospective Adult Research Cohort with Inflammatory Bowel Disease (SPARC IBD).[13] Patients were included if they had an established diagnosis of CD, were ≥18 years of age, and had quiescent disease defined by fecal calprotectin (FCP) < 150 mcg/g, which correlates well with endoscopic healing and suggested as treatment targets by consensus guidelines.[14] Persistent symptoms were defined as mean abdominal pain ≥2 and daily liquid stool frequency ≥4 using the Crohn's Disease Patient-Reported Outcome-2 (CD-PRO2) score, which were suggested as clinically meaningful targets.[14] Exclusion criteria included history of total colectomy or presence of ileostomy or colostomy (see Supplemental Methods for details).

As controls, we included patients from SPARC IBD with active inflammatory CD (**aCD**) defined by FCP >150 mcg/g as well as non-IBD patients recruited at the University of Michigan, including patients meeting Rome IV criteria for diarrhea-predominant irritable bowel syndrome[15] (**IBS-D**) and healthy controls (**HCs**). All patients provided written informed consent prior to enrollment. The Institutional Review Board at University of Michigan gave ethical approval for this work (HUM193179). All authors had access to the study data and reviewed and approved the final manuscript.

### Metagenomic Sequencing

All patients had stool samples available which underwent whole-genome shotgun metagenomic sequencing. Raw whole-genome shotgun reads were then assembled de novo and underwent taxonomic and functional annotation using *geneshot*,[16] a publicly available, open-source Nextflow-based pipeline (please see Supplemental Methods). The complete set of code detailing the metagenomic sequencing pipeline and specific analysis scripts can be found in the GitHub repositories (https://github.com/Golob-Minot/geneshot; https://github.com/aalee1/ibd-sparc.git).

### Statistical Analysis

Continuous variables were compared using t-tests or 1-way analysis of variance and Tukey's test for post-hoc analyses. Fisher's exact test was utilized to compare categorical variables. A 2-tailed $P < .05$ was considered significant after adjustment for multiple comparisons was performed using the Benjamini-Hochberg method (target false discovery rate $q < 0.05$). All analyses were performed using R (version 4.2.1).

Diversity metrics were calculated using the R package {vegan}, including Shannon index as a measure of microbial diversity within-samples ($\alpha$-diversity) as well as Jaccard (presence/absence) and Bray-Curtis (relative abundance) distances for between-sample differences ($\beta$-diversity). Ordination analyses were performed using principal coordinate analysis on Jaccard and/or Bray-Curtis distance matrices at the species level. Differences in $\beta$-diversity metrics between groups were compared using permutational analysis of variance implemented by the "adonis" function in {vegan} using 10,000 permutations. Differential abundance of microbial species was performed using the R package {ANCOM-BC}.[17] Functional profiling of the microbial community was performed using HUMAnN3 based on the KEGG Orthology (KO) database. KO count contingency tables were constructed by presence/absence of genes and compared by Fisher's exact test.

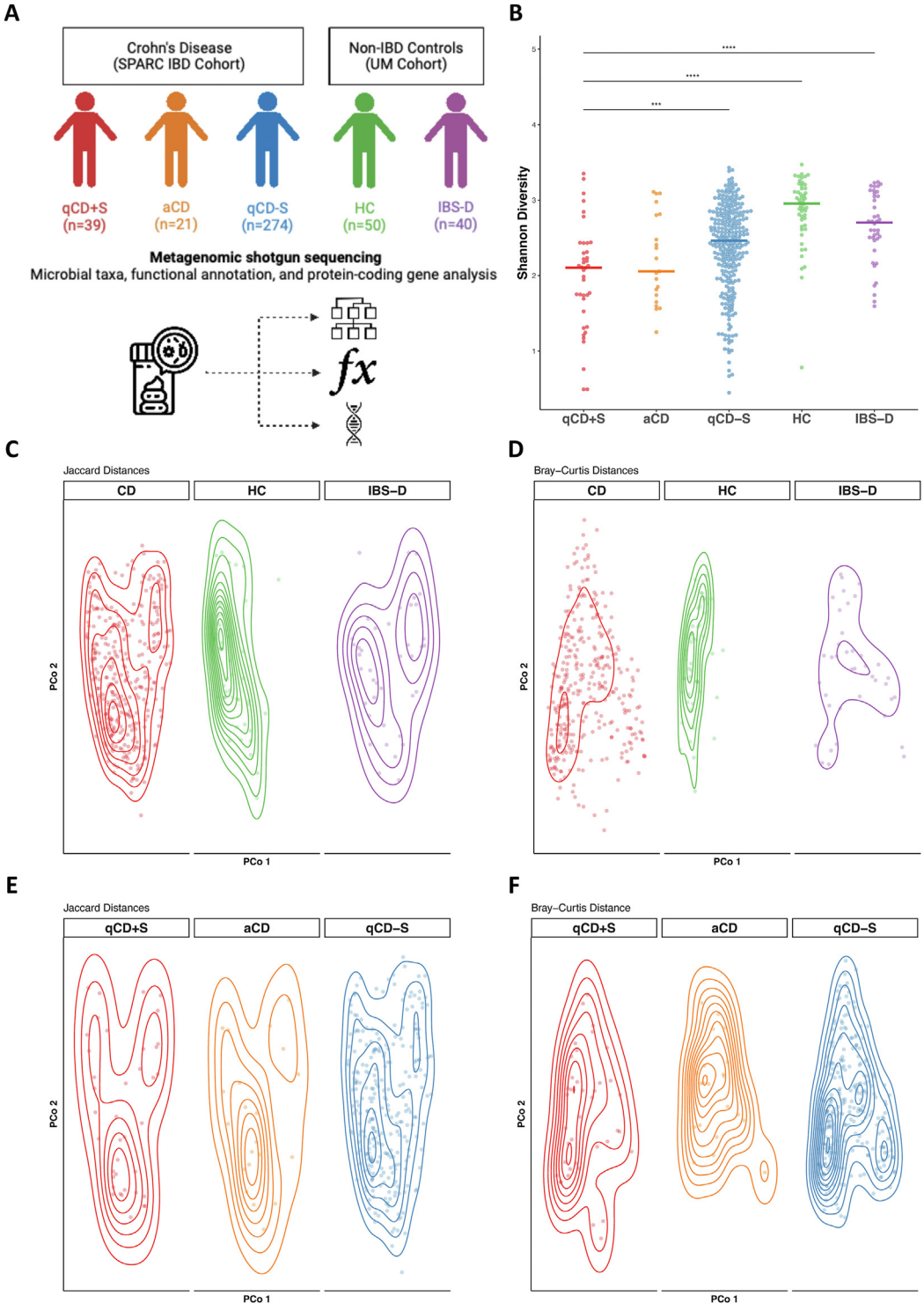



**Figure 1.** The microbiome in qCD + S is similar to aCD but distinct from qCD-S. (A) Stool samples from 39 patients with qCD + S (*red*), 21 with aCD (*orange*), 274 with qCD-S (*blue*), 40 with IBS-D (*green*), and 50 HC (*purple*) underwent WGS metagenomic analysis. (B) Shannon diversity was significantly reduced in qCD + S compared with qCD-S, IBS-D, and HC. Horizontal line represents median value for each group. *** adjusted $P < .001$; **** adjusted $P < .0001$. There were significant differences in microbial community structure at the species level by (C) Jaccard distances ($P = .001$ by PERMANOVA) as well as by (D) Bray-Curtis distances ($P = .00001$ by PERMANOVA) between CD (*red*), HC (*green*), and IBS-D (*purple*). Each dot represents a subject ordinated by PCoA overlaid with 2D-contour plot. Examining only CD subgroups, there were significant differences in microbial community by (E) Jaccard distance between qCD + S (*red*) and qCD-S (*blue*, adjusted $P = .001$), but not compared with aCD (*orange*, adjusted $P = .38$) as well as by (F) Bray-Curtis distances between qCD + S and qCD-S (adjusted $P = .001$), but not compared with aCD (adjusted $P = .22$).

**Table.** Baseline demographics

| Characteristic | SPARC IBD cohort | | | Non-IBD cohort | |
|---|---|---|---|---|---|
| | qCD + S (n = 39) | aCD (n = 21) | qCD-S (n = 274) | IBS-D (n = 40) | HC (n = 50) |
| Age, mean (sd) | 50.3 (14.2) | 40.0 (14.3) | 43.2 (15.1) | 45.6 (18.3) | 46.9 (15.5) |
| Female sex, n (%) | 28 (71.8) | 13 (61.9) | 168 (61.3) | 29 (72.5) | 16 (32.0) |
| Race, n (%) | | | | | |
| White | 31 (91.2) | 15 (78.9) | 225 (92.6) | 37 (92.5) | 37 (74.0) |
| Black | 2 (5.9) | 3 (15.8) | 12 (4.9) | 2 (5.0) | 2 (4.0) |
| Other | 1 (2.9) | 1 (5.3) | 6 (2.5) | 1 (2.5) | 11 (22.0) |
| Ethnicity, n (%) | | | | | |
| Hispanic/Latino | 0 (0.0) | 2 (10.5) | 2 (0.8) | 0 (0.0) | 3 (6.0) |
| Non-Hispanic/Latino | 32 (97.0) | 17 (89.5) | 236 (97.1) | 40 (100.0) | 41 (82.0) |
| Other | 1 (3.0) | 0 (0.0) | 5 (2.1) | 0 (0.0) | 6 (12.0) |
| BMI (kg/m$^2$), mean (sd) | 32.1 (8.0) | 27.2 (6.1) | 28.1 (6.5) | 28.5 (6.8) | 28.0 (4.4) |
| CD Location, n (%) | | | | | |
| Large intestine | 3 (7.7) | 2 (9.5) | 40 (14.6) | | |
| Small and large intestine | 23 (59.0) | 13 (61.9) | 130 (47.4) | | |
| Small intestine | 4 (10.3) | 3 (14.3) | 56 (20.4) | | |
| Other | 9 (23.1) | 3 (14.3) | 48 (17.5) | | |
| Current CD medication, n (%) | | | | | |
| 5-ASA | 2 (5.4) | 2 (9.5) | 34 (12.5) | | |
| Steroids | 8 (21.6) | 11 (52.4) | 20 (7.3) | | |
| Immunomodulators | 11 (29.7) | 5 (23.8) | 88 (32.2) | | |
| Biologics | 28 (75.7) | 15 (71.4) | 189 (69.0) | | |
| None | 5 (13.5) | 2 (9.5) | 32 (11.7) | | |
| Total SES-CD score, mean (sd) | 1.8 (2.0) | 9 (15.6) | 2.3 (4.1) | | |
| Fecal calprotectin (mcg/g stool), mean (sd) | 49.7 (41.8) | 1017.9 (639.1) | 41.3 (41.8) | | |
| History of depression, n (%) | 11 (33.3) | 7 (46.7) | 60 (27.5) | | |
| History of anxiety, n (%) | 15 (45.5) | 6 (40.0) | 58 (26.9) | | |

BMI, body mass index (kg/m$^2$); CD, Crohn's Disease; IBS-D, diarrhea-predominant irritable bowel syndrome; qCD + S, quiescent Crohn's disease with persistent GI, symptoms; qCD-S, quiescent Crohn's disease without persistent GI, symptoms; SES-CD, simple endoscopic score for Crohn's disease; SPARC IBD, Study of a Prospective Adult Research Cohort with Inflammatory Bowel Disease.

Finally, as proof-of-concept to demonstrate the importance of altered gut microbiota and to identify potential biomarkers in qCD + S, we performed random forest classification using the R package {tidymodels} to determine whether microbial species could predict qCD + S vs qCD-S. Data were randomly partitioned with 75% of the data used for model training and 25% held-out for testing. Five-fold cross-validation was employed to estimate model accuracy and to tune hyperparameters. The prediction accuracy of the model was evaluated on the independent test set by calculating the area under the receiver operating characteristic curve. Variable importance scores were extracted to determine the species most important in the model.

# Results

## Baseline Demographics

A total of 424 patients were included, including 39 qCD + S, 274 qCD-S, 21 aCD, 40 IBS-D, and 50 HC (Figure 1A, Table). The median sequencing depth was approximately 25 million read pairs for CD patients and HC and approximately 75 million read pairs for IBS-D patients (Figure A1).

## qCD + S Showed Significant Changes in Microbial Diversity

Patients with qCD + S demonstrated significantly decreased Shannon diversity compared with qCD-S (adjusted $P < .001$); IBS-D and HC (adjusted $P < .0001$ for both); but not compared with aCD (Figure 1B, adjusted $P = .57$).

We also found significant differences in the microbial community between CD, IBS-D, and HC by Jaccard index (Figure 1C, $P = .001$ by permutational analysis of variance) as well as by Bray-Curtis distances (Figure 1D, $P = .00001$ by permutational analysis of variance). Focusing on our primary comparison (ie changes in the microbiome between CD subgroups), the microbial community structure by Jaccard index in qCD + S was significantly different compared with qCD-S (Figure 1E, adjusted $P = .001$), but not compared with aCD (adjusted $P = .38$). Similarly, patients

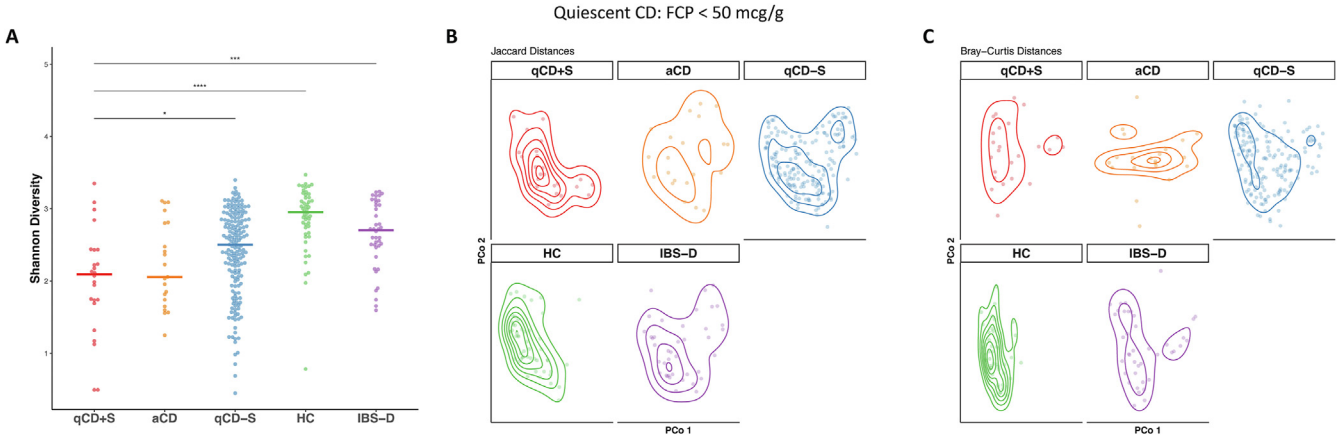



**Figure 2.** Changes in microbial diversity remain despite lowered threshold for quiescent inflammation. These analyses were restricted to quiescent CD patients with FCP <50 mcg/g, including qCD + S (*red*) and qCD-S (*blue*). aCD (*orange*), IBS-D (*purple*), and HC (*green*) were included as before. (A) Patients with qCD + S demonstrated significantly reduced Shannon diversity compared with qCD-S, IBS-D, and HC (adjusted $P = .01$; $P = .0001$; and $P < .0001$, respectively). * adjusted $P < .05$; *** adjusted $P = .0001$; **** adjusted $P < .00001$. There were significant differences in microbial community by (B) Jaccard distances between qCD + S compared with qCD-S, IBS-D, and HC (adjusted $P = .001$ for all groups), but not for aCD (adjusted $P = .39$); as well as by (C) Bray-Curtis distances between qCD + S compared with qCD-S, IBS-D, and HC (adjusted $P = .001$ for all groups), but not for aCD (adjusted $P = .23$).

with qCD + S showed significant differences in microbial community by Bray-Curtis distances compared with qCD-S (adjusted $P = .001$) but not compared with aCD (adjusted $P = .22$, Figure 1F).

## Changes in Microbial Diversity in qCD + S Were Independent From Inflammation

To determine whether our definition for quiescent inflammation may have influenced our findings, we next included only quiescent CD patients with FCP <50 mcg/g, which has the lowest false negative rate for active inflammation in CD.[18] The majority of qCD + S (61.5%) and qCD-S (64.6%) patients met this more stringent criteria for quiescent CD. Overall, our results were not changed by this lower threshold for FCP. Specifically, Shannon index was significantly decreased in qCD + S compared with qCD-S (adjusted $P = .01$), IBS-D (adjusted $P = .0001$), and HC (adjusted $P < .0001$), but not compared with aCD (adjusted $P = .73$, Figure 2A). Microbial community structure by Jaccard index was significantly different in qCD + S compared with qCD-S, IBS-D, and HC (Figure 2B, adjusted $P = .001$ for all comparisons), but not compared with aCD (adjusted $P = .39$). Similarly, the microbial community by Bray-Curtis distances was significantly different in qCD + S compared with qCD-S, IBS-D, HC (adjusted $P = .001$ for all) but not compared with aCD (Figure 2C, adjusted $P = .23$).

## qCD + S Showed Enrichment of Oral Taxa and Depletion of Butyrate/Indole Producers

We next examined which taxa were driving these microbial changes between groups. At the genus level, we identified 42, 52, and 43 taxa that were differentially abundant in qCD + S relative to qCD-S, HC, and IBS-D, respectively, while only 1 taxon was differentially abundant between qCD + S and aCD (Figure A2).

At the species level, 63 species were differentially abundant between qCD + S vs qCD-S (Table A1). Of the species with greatest enrichment (log-fold change>2), a striking number were normal inhabitants of the oral microbiome,[19] including *Streptococcus spp.*, *Rothia dentocariosa*, and *Fusobacterium nucleatum* (Figure 3A). Conversely, the most significantly depleted species (log-fold change $< -2$) included important butyrate and indole producers, including *Eubacterium rectale* and *Faecalibacterium prausnitzii*. Notably, there were no differentially abundant species between qCD + S and aCD ($q > 0.05$).

We next identified microbial species associated with CD-PRO2 score. Although our outcome here was modeled differently (ie, as a continuous rather than a categorical outcome), we found similar species were differentially abundant. Specifically, as CD-PRO2 score increases (indicating worse quality of life scores), we again found enrichment of species that are normal inhabitants of the oral microbiome as well as depletion of important butyrate and indole producers (Figure 3B).

## Functional Metagenome-Related Alterations in qCD + S

We further identified potential functional changes in metagenomes using the KO database. There were significant differences in 152 KO groups between qCD + S and qCD-S, including 91 enriched and 61 depleted (Table A2). Pathways related to cysteine and methionine metabolism (SAM-dependent methyltransferase, homoserine O-acetyltransferase); ATP-binding cassette transporters (phosphonate ATP-binding cassette transporter, ATP-binding cassette

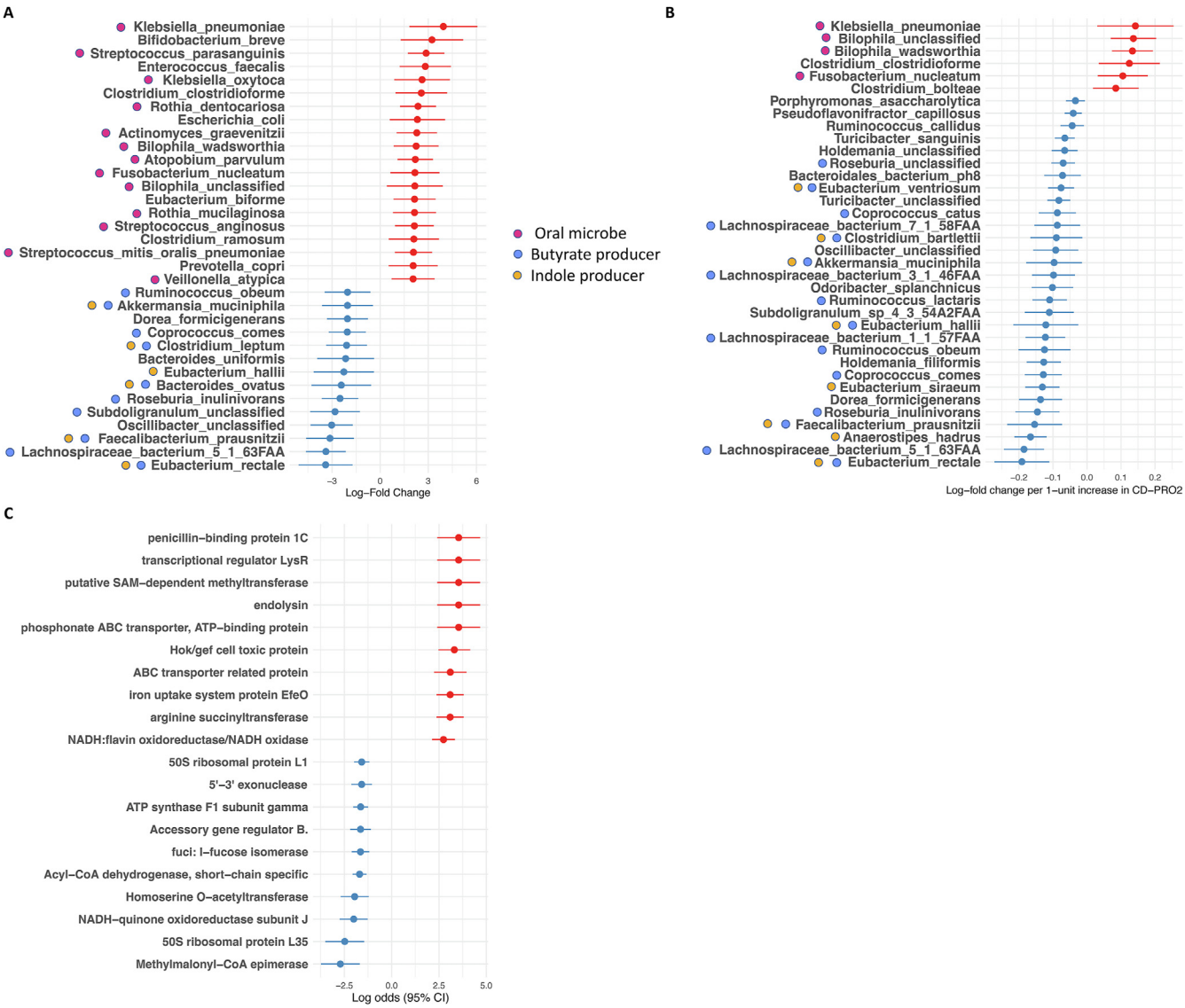



**Figure 3.** qCD + S were enriched with normal inhabitants of the oral microbiome but depleted with important butyrate and indole producers. (A) The most differentially abundant species (absolute log-fold changes [lfc]>2, q < 0.05), including enriched (*red*) and depleted (*blue*) species, in qCD + S relative to qCD-S. Microbial species are annotated as oral microbe (*red* circle), butyrate (*blue* circle), or indole producer (*yellow* circle). (B) Log-fold changes at the species level (q < 0.05) associated with CD-PRO2 score as a continuous outcome. (C) The most dysregulated microbial gene pathways (q < 0.05) by KEGG Orthology, including enriched (*red*) and depleted (*blue*) pathways, in qCD + S compared with qCD-S.

transporter–related protein); fatty acid oxidation (methylmalonyl-CoA epimerase, acyl-CoA dehydrogenase); and iron transport and redox reactions (iron uptake system protein EfeO, NADH:flavin oxidoreductase/NADH oxidase, NADH-quinone oxidoreductase) were among the most differentially represented KO groups in qCD + S compared to qCD-S (Figure 3C).

## Correlation Between PRO Scores, Microbial Diversity and Taxonomy

We next examined the correlation between PRO scores, inflammatory markers, microbial diversity metrics, and species-level taxonomic classification. When including all CD patients, we found that CD-PRO2 scores were positively correlated with inflammatory markers as well as *Escherichia coli* but negatively correlated with diversity metrics (Figure A3A and B).

Next, when examining only quiescent CD patients, we found that PRO measures showed weak positive correlation with inflammatory markers as well as negative correlation with microbial diversity metrics (Figure A3C). Interestingly, we found that CD-PRO2 scores were positively correlated with oral and sulfidogenic microbes as well as negatively correlated with butyrate and indole producers (Figure A3D).

## Bacterial Indole Pathways Were Altered in qCD + S

Given our results showing depletion of important indole producers and to test our a priori hypothesis, we next focused

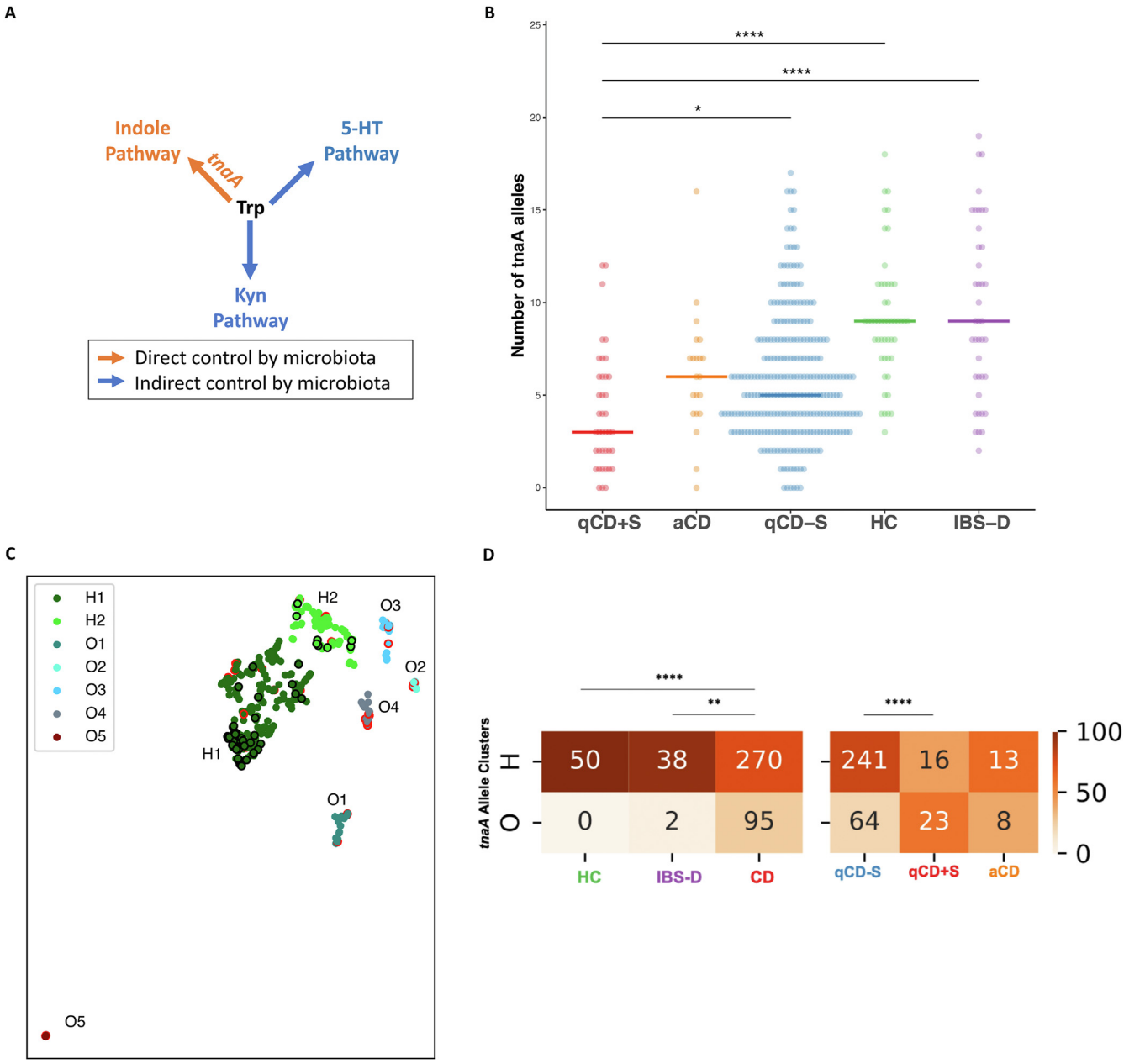



**Figure 4.** qCD + S show reduced and significant allelic variation in *tnaA* microbial genes. (A) Tryptophan (Trp) is catabolized via kynurenine (Kyn), serotonin (5-HT), and indole pathways. While Kyn and 5-HT pathways are mediated by both host and microbial enzymes, indole pathways are solely mediated by microbial enzymes, including tryptophanase (*tnaA*). (B) The number of *tnaA* alleles was significantly reduced in qCD + S compared with qCD-S, HC, and IBS-D, but not for aCD. (C) Ordination of *tnaA* alleles by Jaccard distance using UMAP demonstrated 7 distinct clusters. HC (*black* outline) universally clustered into 2 clusters (labeled as 'Healthy' H1 and H2 clusters) while CD patients, particularly qCD + S (*red* outline), showed a greater distribution outside the 'Healthy' clusters (labeled as 'Other' O1-O5 clusters). (D) The proportion of CD patients clustering into the 'Healthy' (H) clusters was significantly different compared with HCs and IBS-D patients. These differences in clustering were mainly driven by qCD + S compared with qCD-S and aCD. Numbers represent the number of patients from each group within each cluster while shading represents proportion of patients within each cluster. ** adjusted $P < .005$; **** adjusted $P < .0001$.

on tryptophan pathways in qCD + S. Tryptophan is catabolized by 3 major pathways, including kynurenine (Kyn), serotonin (5-HT), and indole (Figure 4A). While Trp metabolism to Kyn and 5-HT are mediated by both host and microbial enzymes, microbial enzymes are solely responsible for metabolism of tryptophan to indole derivatives.[12] Indole is the most abundant of the microbial-derived Trp catabolites in the human gut with concentrations >500-fold higher compared with the next most common indole catabolite, indoleacetic acid.[20] Given the predominance of indole in the human gut as

well as indole's role in modulating intestinal barrier function, known to be impaired in qCD + S,[6,7] we specifically focused our analyses on the bacterial *tnaA* gene encoding the enzyme tryptophanase, which catalyzes the conversion of Trp to indole. Microbial *tnaA* alleles were significantly decreased in qCD + S compared with qCD-S (adjusted $P = .02$), HC, and IBS-D patients (adjusted $P < .0001$ for both) but not compared with aCD (adjusted $P = .19$, Figure 4B).

As a complementary method to taxonomy-based approaches, we next considered how individual *tnaA* alleles related to CD phenotypes. Given the multicollinearity between *tnaA* alleles as well as to further determine whether the diversity and distribution of individual *tnaA* alleles may relate to phenotype, we employed an ordination-and-clustering based approach focused on presence/absence of *tnaA* alleles, rather than small changes in relative abundance. We performed uniform manifold approximation and projection, a nonlinear dimensionality reduction technique, combined with {HDBSCAN}, an unsupervised clustering algorithm, on Jaccard distances of *tnaA* alleles. We noted that all *tnaA* alleles from healthy adults fell into 1 of 2 very closely related clusters that we labeled "healthy" (H) clusters, numbered H1-H2. There were 4 additional clusters identified that we labeled "other (O)" clusters, and numbered O1-O4 (Figure 4C). There were significantly decreased number of CD patients (74.0%) clustering into "healthy" (H) clusters compared with HC (100%; adjusted $P < .0001$) and IBS-D (95.0%; adjusted $P = .003$, Figure 4D). Focusing on CD subgroups, the proportion of qCD + S clustering into "healthy" (H) clusters was significantly lower compared with qCD-S (41.0% vs 79.0%, adjusted $P < .0001$, Figure 4D) but not compared to aCD (61.9%; adjusted $P = .20$). As the combination of uniform manifold approximation and projection with HDBSCAN has been shown to improve identification of biologically meaningful clusters,[21] these results suggest a distinct *tnaA* allelic variation in qCD + S.

### Microbiome Predicted qCD + S

As a complementary approach to our compositional analysis and to explore potential biomarkers for qCD + S, we next attempted to predict qCD + S vs qCD-S based on species-level taxonomic composition. After splitting the available data 75%–25% into a train/test set, we then trained a Random Forest Classifier model. The prediction of the model showed good accuracy by 5-fold cross-validation (area under the receiver operating characteristic curve 0.73) on the training set as well as when validated using the held-out test set (area under the receiver operating characteristic curve 0.72; bootstrapped 95% confidence interval: 0.58–0.86) (Figure 5A). Reassuringly, many of the species that were important to the model by variable importance scores (Figure 5B) were also found to be differentially abundant by {ANCOM-BC} (Figure 3A).

## Discussion

The mechanisms underpinning qCD + S are poorly understood despite being a prevalent and morbid condition in CD. In this multicenter study, we demonstrated that patients with qCD + S showed reduced microbial diversity, changes in microbial community structure and composition, and potential microbial function relative to qCD-S, IBS-D, and HC. Speaking to potential mechanisms, qCD + S demonstrated a striking enrichment of species normally found in the oral microbiome and potential functional changes related to sulfur metabolism paired with depletion of important butyrate and indole producers. Additionally, qCD + S showed potential changes in microbial indole pathways, including depletion of microbial *tnaA* genes and evidence of *tnaA* allelic variation. Finally, we showed that microbial species were able to predict presence of qCD + S with reasonable accuracy.

A novel finding in our study is that although qCD + S share similar microbial diversity, community structure, and membership with aCD, there were significant differences in these microbial indices compared with qCD-S. Indeed, while 63 species were differentially abundant between qCD + S vs qCD-S, no species were differentially abundant between qCD + S and aCD. If we consider variation in the microbial community as a marker of host well-being that spans the continuum from health to disease, these findings suggest that the microbiome in qCD + S is shifted towards an inflammatory phenotype while the community in qCD-S may have reconstituted more towards healthy controls.

While prior studies have consistently demonstrated intestinal epithelial barrier dysfunction in qCD + S,[6,7] the mechanisms are largely unknown.[8] We identified depletion of multiple bacterial species important in butyrate and indole production in qCD + S.[12] As butyrate and indole metabolites are important regulators of epithelial barrier integrity,[12] these results suggest that altered gut microbiota may mediate epithelial barrier disruption in qCD + S.

In line with these findings, we identified significant disruption in microbial-tryptophan pathways in qCD + S compared with qCD-S, but not compared with aCD. Specifically, we demonstrated that qCD + S had significant depletion of *tnaA* alleles (the microbial gene responsible for catabolism of tryptophan into indoles) compared with qCD-S. Although *tnaA* is one of several bacterial genes involved in Trp metabolism, indole is by far the most abundant of the microbially derived Trp metabolites and known to influence intestinal barrier function.[20] Furthermore, the specific *tnaA* alleles present were distinct in qCD + S compared to qCD-S. While most patients with qCD-S clustered with HC, qCD + S showed significantly higher membership outside of this healthy cluster. As this ordination and clustering approach using uniform manifold approximation and projection with HDBSCAN has been shown to improve identification of clusters with shared biological functions and distinct clinical phenotypes,[21] this allelic variation in bacterial *tnaA* genes suggest significant variability in the microbial community's ability to compete for tryptophan and produce indole. Future studies are required to determine whether disruption in indole metabolism may contribute to intestinal barrier dysfunction in qCD + S.

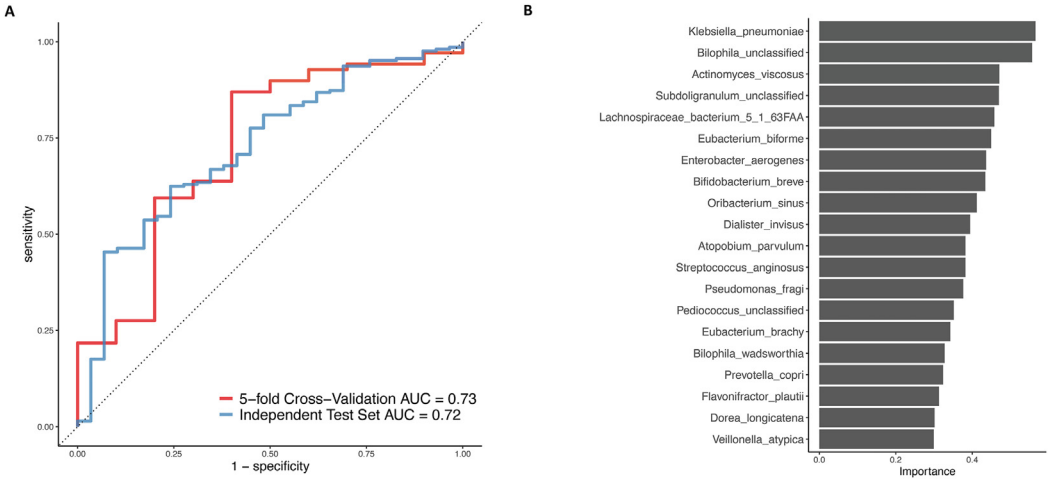



**Figure 5.** Microbial species accurately predicted qCD + S. (A) After randomly splitting the data into training (75%) and test (25%) set, a random forest model with 5-fold cross-validation was trained on species abundance. The model showed good accuracy for the training data (AUC = 0.73, *red line*) as well as when validated on the independent test set (AUC = 0.72, *blue line*). (B) The top twenty species by variable importance scores are illustrated for the random forest model. area under the receiver operating characteristic curve.

Another novel finding in our study is the enrichment of bacterial species that are normal inhabitants of the oral microbiome. Normally, the oral microbiome is distinct from the gut microbiome in healthy adults.[22] However, a recent study demonstrated that the oral and gut microbiome were more similar in CD patients, most of whom were in clinical remission, compared with HC,[23] and thus suggests gut colonization by oral bacteria in CD. Additionally, the presence of periodontitis increased future risk for flare in CD symptoms as measured by the short Crohn's Disease Activity Index (sCDAI). As the sCDAI is only weakly correlated with inflammation,[24] this study may also have been demonstrating the relationship between oral-gut translocation of microbes and persistent symptoms in quiescent CD. Although the functional consequences of oral-gut translocation in qCD + S are unknown, cysteine and methionine metabolism, ATP transport, and redox reactions were among the most highly dysregulated pathways in qCD + S. Cysteine and methionine are 2 primary sulfur-containing amino acids, while many oral microbes are known to be producers of hydrogen sulfide ($H_2S$).[25] As $H_2S$ is an important regulator of mitochondrial function, oxidative stress, and intestinal permeability,[26] these data are suggestive of $H_2S$ as a novel pathway in qCD + S. Furthermore, it is currently difficult to predict which patients may develop qCD + S. Our predictive modeling approach suggested that oral and sulfidogenic microbes may be important biomarkers. Future studies should focus on whether these microbial species may be used as biomarkers to predict CD patients who are at high risk for developing qCD + S.

There are several strengths of our work. First, to improve upon prior work that were small single-center studies, we utilized data from the multicenter, prospective SPARC IBD cohort. Given the variability of the microbiome in CD,[27] this strategy of sampling from a geographically, racially, and ethnically diverse cohort across the US helped to generalize our findings. Furthermore, patients enrolled in SPARC IBD undergo comprehensive phenotyping, while sample processing and sequencing are highly standardized using a single reference laboratory, which reduces bias and improves reproducibility.[13] In addition, the inclusion of data from aCD and non-IBD patients allowed us to contextualize how the microbiome varies from active inflammation to health. Additionally, our results consistently suggested the importance of oral and sulfidogenic microbes as well as depletion of butyrogenic and indole producing microbes in qCD + S regardless of the statistical approach (eg, ANCOM-BC, correlation analysis, or random forest). Finally, while common metagenomic approaches are limited by the robustness of reference databases, our bioinformatics approach focused on microbial protein-coding genes that are assembled de novo and thus do not have these same constraints. Furthermore, this approach allowed us to evaluate our a priori hypothesis on bacterial *tnaA* genes and infer the metagenomic potential of the microbial community.

However, there are limitations to our work as well. First, only a subset of patients had endoscopic evidence of quiescent disease, and thus, we cannot completely exclude the possibility that our results were influenced by inflammation. Furthermore, we cannot exclude the possibility of deeper bowel wall inflammation even with lower FCP levels and/or mucosal healing on colonoscopy.[28] Although prior studies have demonstrated good-excellent ability with FCP to discriminate between endoscopically active and quiescent CD, the relationship between FCP and endoscopy is imperfect, especially in isolated ileal disease.[14] However, as our results remained robust even when we restricted our analysis to patients with FCP <50 mcg/g, which has a negative predictive value of 94% for excluding

inflammation,[18] this suggests that the microbiome changes described here are largely independent from inflammation. Second, while these results support an association between microbial changes and persistent symptoms in quiescent CD, we cannot make any firm conclusions on causation here. Similarly, we do not have longitudinal data to suggest directionality between microbial changes and persistent symptoms. Third, as CD and non-CD patients were sequenced at different laboratories, we cannot exclude the possibility of technical bias in our analyses. However, our primary aim was to compare the microbiome across CD patients, which were all sequenced at a single reference lab using identical DNA extraction and sequencing protocols. Thus, the possibility of technical bias would not be expected to affect our primary results. Furthermore, all reads from both CD and non-CD patients were pre-processed and processed computationally in an identical manner, with the same pipeline, hyperparameters, and reference libraries to reduce the risk of computational-induced artifacts. Finally, although diet is known to be a major influence on the microbiome, we did not have dietary information on CD patients.

## Conclusion

In conclusion, we have identified quiescent CD patients with persistent symptoms have widespread changes in their intestinal microbiota similar to the microbial community found in active CD but quite distinct from quiescent CD without persistent symptoms. These proof-of-concept findings suggest an altered microbiome may mediate intestinal barrier dysfunction, while enrichment of oral microbes and potential functional changes related to sulfur and tryptophan metabolism may be novel pathways in qCD + S. While our findings require validation, future work may determine whether microbiome-based therapies show promise as novel therapeutic approaches for persistent symptoms in quiescent CD.

## Supplementary Materials

Material associated with this article can be found in the online version at https://doi.org/10.1016/j.gastha.2023.11.005.

## References


1. Halpin SJ, Ford AC. Prevalence of symptoms meeting criteria for irritable bowel syndrome in inflammatory bowel disease: systematic review and meta-analysis. Am J Gastroenterol 2012;107:1474–1482.
2. Gracie DJ, Williams CJM, Sood R, et al. Negative effects on psychological health and quality of life of genuine irritable bowel syndrome–type symptoms in patients with inflammatory bowel disease. Clin Gastroenterol Hepatol 2017;15:376–384.e5.
3. Anderson A, Click B, Ramos-Rivers C, et al. The association between sustained poor quality of life and future opioid use in inflammatory bowel disease. Inflamm Bowel Dis 2018;24:1380–1388.
4. Lichtenstein GR, Feagan BG, Cohen RD, et al. Serious infection and mortality in patients with Crohn's disease: more than 5 years of follow-up in the TREAT™ registry. Am J Gastroenterol 2012;107:1409–1422.
5. Limsrivilai J, Stidham RW, Govani SM, et al. Factors that predict high health care utilization and costs for patients with inflammatory bowel diseases. Clin Gastroenterol Hepatol 2017;15:385–392.e2.
6. Chang J, Leong RW, Wasinger VC, et al. Impaired intestinal permeability contributes to ongoing bowel symptoms in patients with inflammatory bowel disease and mucosal healing. Gastroenterology 2017;153:723–731.e1.
7. Vivinus-Nébot M, Frin-Mathy G, Bzioueche H, et al. Functional bowel symptoms in quiescent inflammatory bowel diseases: role of epithelial barrier disruption and low-grade inflammation. Gut 2014;63:744–752.
8. Henriksen M, Høivik ML, Jelsness-Jørgensen L-P, et al. Irritable bowel-like symptoms in ulcerative colitis are as common in patients in deep remission as in inflammation: results from a population-based study [the IBSEN study]. J Crohns Colitis 2018;12:389–393.
9. Boland K, Bedrani L, Turpin W, et al. Persistent diarrhea in patients with Crohn's disease after mucosal healing is associated with lower diversity of the intestinal microbiome and increased dysbiosis. Clin Gastroenterol Hepatol 2021;19:296–304.e3.
10. Humbel F, Rieder JH, Franc Y, et al. Association of alterations in intestinal microbiota with impaired psychological function in patients with inflammatory bowel diseases in remission. Clin Gastroenterol Hepatol 2020;18:2019–2029.e11.
11. Borren NZ, Plichta D, Joshi AD, et al. Alterations in fecal microbiomes and serum metabolomes of fatigued patients with quiescent inflammatory bowel diseases. Clin Gastroenterol Hepatol 2021;19:519–527.e5.
12. Agus A, Clément K, Sokol H. Gut microbiota-derived metabolites as central regulators in metabolic disorders. Gut 2021;70:1174–1182.
13. Raffals LE, Saha S, Bewtra M, et al. The development and initial findings of A study of a prospective adult research cohort with inflammatory bowel disease (SPARC IBD). Inflamm Bowel Dis 2022;28:192–199.
14. Turner D, Ricciuto A, Lewis A, et al. STRIDE-II: an update on the selecting therapeutic targets in inflammatory bowel disease (STRIDE) initiative of the international organization for the study of IBD (IOIBD): determining therapeutic goals for treat-to-target strategies in IBD. Gastroenterology 2021;160:1570–1583.
15. Mearin F, Lacy BE, Chang L, et al. Bowel disorders. Gastroenterology 2016;150:1393–1407.
16. Minot SS, Barry KC, Kasman C, et al. geneshot: gene-level metagenomics identifies genome islands associated with immunotherapy response. Genome Biol 2021;22:135.
17. Lin H, Peddada SD. Analysis of compositions of microbiomes with bias correction. Nat Commun 2020;11:3514.
18. Kopylov U, Yung DE, Engel T, et al. Fecal calprotectin for the prediction of small-bowel Crohn's disease by capsule endoscopy: a systematic review and meta-

analysis. Eur J Gastroenterol Hepatol 2016; 28:1137–1144.
19. Structure, function and diversity of the healthy human microbiome. Nature 2012;486:207–214.
20. Roager HM, Licht TR. Microbial tryptophan catabolites in health and disease. Nat Commun 2018;9:3294.
21. Yang Y, Sun H, Zhang Y, et al. Dimensionality reduction by UMAP reinforces sample heterogeneity analysis in bulk transcriptomic data. Cell Rep 2021;36:109442.
22. Rashidi A, Ebadi M, Weisdorf DJ, et al. No evidence for colonization of oral bacteria in the distal gut in healthy adults. Proc Natl Acad Sci U S A 2021;118: e2114152118.
23. Imai J, Ichikawa H, Kitamoto S, et al. A potential pathogenic association between periodontal disease and Crohn's disease. JCI Insight 2021;6:e148543.
24. Jones J, Loftus EV, Panaccione R, et al. Relationships between disease activity and serum and fecal biomarkers in patients with Crohn's disease. Clin Gastroenterol Hepatol 2008;6:1218–1224.
25. Wolf PG, Cowley ES, Breister A, et al. Diversity and distribution of sulfur metabolic genes in the human gut microbiome and their association with colorectal cancer. Microbiome 2022;10:1–16.
26. Murphy B, Bhattacharya R, Mukherjee P. Hydrogen sulfide signaling in mitochondria and disease. FASEB J 2019;33:13098–13125.
27. Clooney AG, Eckenberger J, Laserna-Mendieta E, et al. Ranking microbiome variance in inflammatory bowel disease: a large longitudinal intercontinental study. Gut 2021;70:499–510.
28. Laharie D, D'Haens G, Nachury M, et al. Steroid-free deep remission at one year does not prevent Crohn's disease progression: long-term data from the TAILORIX trial. Clin Gastroenterol Hepatol 2022;20:2074–2082.

**Received October 10, 2023. Accepted November 9, 2023.**

**Correspondence:**
Address correspondence to: Allen A. Lee, MD, MS, Division of Gastroenterology, Department of Internal Medicine, University of Michigan Medical School, 3912 Taubman Center, 1500 E. Medical Center Dr., SPC 5362, Ann Arbor, Michigan 48109. e-mail: allenlee@umich.edu.

**Authors' Contributions:**
Study concept and design: Jonathan Golob, Krishna Rao, Shrinivas Bishu, Allen A. Lee. Analysis or interpretation of data: All authors. Drafting of the manuscript: All authors. Figures: Jonathan Golob, Krishna Rao, Shrinivas Bishu, Allen A. Lee. Critical revision of the manuscript: All authors. Final approval: All authors.

**Conflicts of Interest:**
The authors disclose no conflicts.

**Funding:**
The results published here are in whole or part based on data obtained from the IBD Plexus program of the Crohn's & Colitis Foundation. This study was also supported by grants from The Leona M. and Harry B. Helmsley Charitable Trust (to AAL) and the National Institutes of Health grants DK124567 (to AAL), HS027431 (to KR), and DK123403 (to SB).

**Ethical Statement:**
The corresponding author, on behalf of all authors, jointly and severally, certifies that their institution has approved the protocol for any investigation involving humans or animals and that all experimentation was conducted in conformity with ethical and humane principles of research. The institutional review board at UM gave ethical approval for this work (HUM193179). All authors had access to the study data and reviewed and approved the final manuscript.

**Data Transparency Statement:**
Sequencing data is available by request through the IBD Plexus (please see https://www.crohnscolitisfoundation.org/research/grants-fellowships/ibd-plexus for full details). The complete set of code detailing the metagenomic sequencing pipeline and specific analysis scripts can be found in the GitHub repositories (https://github.com/Golob-Minot/geneshot; https://github.com/aalee1/ibd-sparc.git).

**Reporting Guidelines:**
STROBE.

**Preprint:**
https://doi.org/10.1101/2023.05.16.23290065.