Thank you to the editors and reviewers for your time and consideration evaluating our manuscript, “Fecal microbiota signatures are associated with response to Ustekinumab therapy among Crohn's Disease patients” (mBio02120-17). The reviewers raised excellent points and have helped make the manuscript stronger. Attached you will find our responses to the reviewer comments.

Response to Reviewer Comments.

Reviewer #1 (Comments for the Author):
1. A table or description of the numbers of fecal samples collected at each time point would be beneficial. Figure 4 shows a significant difference between week 0 and 22 for treated responders (n=18). There are no significant differences in the Placebo groups but the numbers are a lot lower (n=6/8). I assume that these were the maximum possible but it is not obvious.

These numbers are presented in the methods section of the original version (lines 327-330) and for the primary and secondary endpoints in Table 2, but we clarified that subjects provided a single sample at each time point and moved the sample numbers to the study design section in the revised manuscript (lines 307-317).

The assumption that this is the maximum possible for Figure 4 is correct, as only subjects who provided a sample at every time point were included for this part of our analysis (lines 190-193). This greatly reduced the number of subjects/samples we could include and resulted in the low statistical power for the Placebo group and change over time.

Reviewer #2 (Comments for the Author):
2. According to Table 1, it appears that the treated vs. placebo cohorts are different regarding sex, corticosteroid use, and even tissue involvement. Are these groups statistically different for these clinical variables? And if so how might that influence the observations reported?

The placebo and treated groups were not statistically different by sex ($p=0.31$) or corticosteroid use ($p=0.06$) using a Wilcoxon rank sum test. Additionally these groups were not statistically different by tissue involvement following a Kruskal-Wallis and multiple comparisons test ($p>0.05$). A footnote was added to Table 1 to address this in the revised manuscript (page 17).

3. Please include details on inclusion/exclusion criteria for the study.

The original study publication describes the inclusion/exclusion criteria in detail within the ‘Protocol’ of the supplementary materials. (Reference #37. Sandborn WJ et al. 2012. Ustekinumab induction and maintenance therapy in refractory Crohn’s disease. N Engl J Med 367:1519–28. http://www.nejm.org/doi/suppl/10.1056/NEJMoa1203572/suppl_file/nejmoa1203572_protocol.pdf). We have clarified this in the revised version and briefly describe the inclusion criteria (lines 303-307).

4. How were concomitant medications handled? Obviously some were on corticosteroids, but what about antibiotics and other systemic meds? Concomitant inflammatory and/or metabolic disease? (For example psoriasis.)

See above response – the 'Methods/Patients' section of the original study publication describes how concomitant medications were handled. As for concomitant inflammatory and metabolic diseases/conditions, patients were not excluded based on these.

5. Did subjects self-collect stool? And if so how was it ensured the sample was not degraded prior to extraction and sequencing?

Stool samples were collected by the patients at home, kept refrigerated for no more than 24h, and then brought to the clinical sites and frozen. We have clarified this in the revised manuscript (lines 317-319)

6. It is surprising that so many of the Placebo subjects experienced response or remission. Do those achieving response/remission in treated and placebo groups have similar characteristics?

It is surprising, but our data reflect observed placebo effects, as described in "Placebo effects and their determinants in gastrointestinal disorders." Nature Reviews Gastroenterology & Hepatology. 12, 472–485 (2015). doi:10.1038/nrgastro.2015.117.

> "According to the listed meta-analyses in Table 1, the pooled placebo response ranges between 20–35% in IBD, and 20–40% in IBS, functional dyspepsia, gastric and duodenal ulcer and GERD, with marginal differences between functional and somatic disorders (see Table 1).
> " https://www.nature.com/articles/nrgastro.2015.117/tables/1

Additionally the original study publication (reference #37) shows a statistical difference in the treated vs. untreated populations in terms of response/remission. As a whole our treated vs. untreated groups were similar at the outset, with respect to the microbiome, and we did not observe meaningful similarities between treatment and placebo groups that correlated to responding/remitting.

7. This raises the point that it would be useful to generate a simple figure or table that summarizes the results. It is difficult to keep track of the comparisons made and the commonalities revealed (for example, Faecalibacterium and Escherichia/Shigella

This point is well taken. A table was added (Table 3, page 19) and referred to in the discussion section of the revised manuscripts (line 237).

8. It is stated that CDAI weighs stool frequency, abdominal pain, general well being, weight change, hematocrit, opioid usage among others to determine the severity

score. How did the individual characteristics of the CDAI score correlate with microbiome? For example can you split out what was the most dominant factor contributing to this score in each subject and further stratify the cohort based on this?

We saw no significant associations between quantitative CDAI sub scores and the microbiome, with the exception of loose stool frequency. Qualitative sub scores were not considered due to their subjectivity. This was addressed in the Discussion section of the original version (lines 264-268), but is clarified in the revised manuscript (lines 266-270),