nature research

Corresponding author(s): Dr. Andrew J. Macpherson

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Please do not complete any field with "not applicable" or n/a. Refer to the help text for what text to use if an item is not relevant to your study.
For final submission: please carefully check your responses for accuracy; you will not be able to make changes later.

Experimental design

1. Sample size

Describe how sample size was determined.

Both in Swiss IBD Cohort (Cohort 1) and Bern Uni. Hosp. Cohort (Cohort 2), we have extracted the biopsy samples of patients who were sampled during the colonoscopy for 5 different intestinal locations (ileum, right colon, left colon, transverse colon and rectum). No sample size calculation was performed. This was a meta-analysis that used all available samples of participants with biopsies from 5 different locations.

2. Data exclusions

Describe any data exclusions.

In this study we have collected the biopsy samples from 5 different intestinal locations per patient. We processed all samples however, in some cases we didn't get enough DNA that was sufficient for 16S rRNA PCR samples and hence, those samples were excluded. After sequencing, samples with more than 4500 high quality reads were used for further analysis. We have established the cut-off criteria based on a rarefaction analysis and phylogenetic resolution that we obtained with alpha diversity analysis.

3. Replication

Describe the measures taken to verify the reproducibility of the experimental findings.

We have run this study with two independent cohorts. Cohort 1 is Swiss IBD cohort that exists more than years. We have first obtained samples from this cohort and sequenced for 16S rRNA microbial profile. Then, as replica set we have the Cohort 2 that consists of newly recruited IBD patients and non-IBD subjects in the framework of the human intestinal community project of Bern University Hospital. We have analyzed the data for Cohort 2 with the settings we established for Cohort 1 and in our manuscript we reported both cohort results side by side. The experimental findings in Cohort 1 were reliably reproduced through repeated experiments using Cohort 2.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

IBD Patients were assigned with unique ID number according to their recruitment time and we processed/sequenced the sample by increasing number order and hence, patient with different disease status were randomized. Randomization is all dependent on the recruitment time of the patients.

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

Biopsy samples were assigned with unique patient ID number and hence, until we sequence biopsy samples for microbial profile we didn't have information of the clinical metadata information of the patients.

Note: all in vivo studies must report how sample size was determined and whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐ A statement indicating how many times each experiment was replicated
☐	☒ The statistical test(s) used and whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒ Test values indicating whether an effect is present *Provide confidence intervals or give results of significance tests (e.g. P values) as exact values whenever appropriate and with effect sizes noted.*
☐	☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒ Clearly defined error bars in all relevant figure captions (with explicit mention of central tendency and variation)

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.

We initially analyzed the 16s rRNA data using QIIME v1.9.1. OTU picking step assigns OTUs by clustering sequences based on a user-defined similarity threshold, namely UCLUST (Edgar, RC 2010) that creates "seeds" of sequences which generate clusters based on percent identity. The otu_table.biom file, mapping file (prepared using Microsoft Excel 2016) and rep_set.tree file were loaded into phyloseq package in R version 3.3.0 for further analysis. Plots were generated using phyloseq object or GraphPad Prism v7. Taxonomy analysis were performed in Maaslin in R and all-to-all hierarchical clustering analysis was performed using HAllA (https://huttenhower.sph.harvard.edu/halla) from C. Huttenhower's group. Microbial association network construction within disease status were generated using script provided in manuscript Williams et. al., Demonstrating microbial co-occurrence pattern analyses within and between ecosystems. Frontiers in Microbiology 5. and network plots were generated using Cytoscape v3.5.1. All reaction level analysis were performed in MATLAB (R2018a, MathWorks)

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a third party.

All 16S rRNA sequencing data / Clinical metadata of patients are readily available from the authors.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

No antibodies were used.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

No eukaryotic cell lines were used.

b. Describe the method of cell line authentication used.

No eukaryotic cell lines were used.

c. Report whether the cell lines were tested for mycoplasma contamination.

No eukaryotic cell lines were used.

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.

No eukaryotic cell lines were used.

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide all relevant details on animals and/or animal-derived materials used in the study.

Study did not involve any mouse models.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

Biopsy samples in this study were collected either a) in the Swiss IBD cohort study or b) newly recruited IBD patients and non-IBD subjects in the framework of the human intestinal community project of Bern University Hospital. Non-IBD biopsy samples were collected from subjects that were registered for a screening ileo-colonoscopy without any gastrointestinal symptoms, no suspected functional intestinal symptoms and where all other work-up was negative. Age range of the participants into this study is between 20-75 years and the gender (% female) value is CD: 56%, UC: 60% in Cohort 1 and CD: 42%, UC: 21% , Non-IBD: 53% in Cohort 2. More than 90% of the participants are Caucasian and mean disease duration in these patients in each cohort is around 10 years. The descriptions of Inclusion criteria of human participants have been provided in the "Cohorts and study design" method section. Briefly, inclusion criteria for Swiss IBD cohort patients were on established Lennard-Jones criteria, confirmed by radiological, endoscopic and histological findings. Additional characteristics, including disease activity, surgery status, disease location, remission status and therapy are detailed in Suppl. Table 1 for Cohort 1 and for Suppl. Table 2 for Cohort 2.

Licensed gastroenterologists collected biopsy samples and clinical data of patients identified as CD or UC and non-IBD subjects. The Swiss IBD Cohort study was approved by all local ethical committees of the participating hospitals: The Bern Human Intestinal Community project was approved by the Bern Cantonal Ethics Commission (Ref: KEK-BE: 251/14 and 336/14). This statement is included in the methods section.