

Peer Review Information

Journal: Nature Methods

Manuscript Title: Microbial-enrichment method enables high-throughput metagenomic characterization from host-rich samples

Corresponding author name(s): Rustem F. Ismagilov

Editorial Notes:

Reviewer Comments & Decisions:

Decision Letter, initial version:
--

26th Oct 2022

Dear Rustem,

Thank you for your inquiry about submitting your manuscript, "Microbial-enrichment method enables high-throughput metagenomic characterization from host-rich samples" to Nature Methods. The paper sounds like it may be of interest, and should fit the scope of the journal.

Of course, it is very difficult to judge a paper based only on the limited information available in a presubmission inquiry. At this stage, we are unsure about whether the proposed MEM treatment represents a substantially methodology development. Therefore I am sure you understand that we cannot promise to send your paper out for peer review and must read it in its entirety before deciding if this would be suitable.

We strongly encourage you to include data to validate method performance and demonstrate its advantage over existing tools and its general applicability.

You will find our Guide to Authors at <http://www.nature.com/naturemethods> to assist you in preparing your manuscript. However, it is not necessary at this stage to spend major effort adhering to our detailed formatting instructions.

Thank you for your interest in Nature Methods.

Best regards,
Lei

Lei Tang, Ph.D.
Senior Editor
Nature Methods

Author Rebuttal to Initial comments

[There is no Author Rebuttal letter at this stage.]

Decision Letter, first revision:

1st Mar 2023

Dear Professor Ismagilov,

Your Article, "Microbial-enrichment method enables high-throughput metagenomic characterization from host-rich samples", has now been seen by 3 reviewers. As you will see from their comments below, although the reviewers find your work of considerable potential interest, they have raised a number of concerns. We are interested in the possibility of publishing your paper in Nature Methods, but would like to consider your response to these concerns before we reach a final decision on publication.

We therefore invite you to revise your manuscript. The required data and revisions include:

- resolving the technical issues raised by reviewers
- providing software and code available for testing by the reviewers (please see below the code availability section)

We are committed to providing a fair and constructive peer-review process. Do not hesitate to contact us if there are specific requests from the reviewers that you believe are technically impossible or unlikely to yield a meaningful outcome.

When revising your paper:

- * include a point-by-point response to the reviewers and to any editorial suggestions
- * please underline/highlight any additions to the text or areas with other significant changes to facilitate review of the revised manuscript
- * address the points listed described below to conform to our open science requirements
- * ensure it complies with our general format requirements as set out in our guide to authors at

www.nature.com/naturemethods

* resubmit all the necessary files electronically by using the link below to access your home page

[REDACTED]

Note: This URL links to your confidential home page and associated information about manuscripts you may have submitted, or that you are reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage.

We hope to receive your revised paper within 14 weeks. If you cannot send it within this time, please let us know. In this event, we will still be happy to reconsider your paper at a later date so long as nothing similar has been accepted for publication at Nature Methods or published elsewhere.

CODE AVAILABILITY

To facilitate software and code availability, we offer a service in partnership with [Code Ocean](https://codeocean.com/) to enable authors to share fully-functional and executable code accompanying their articles and to facilitate peer review of code by the reviewers.

The use of the Code Ocean platform for peer review of code associated with this paper will be under the same confidentiality and anonymity agreements as the rest of manuscript materials.

Code Ocean is a cloud-based reproducibility platform where authors upload code and data and configure the necessary computational environment for reproduction. The code, data, metadata, and computational environment -- called a 'compute capsule' -- can then be accessed by reviewers in an anonymous fashion, and upon publication, provided to readers via a link from the article. Code Ocean supports all open source programming languages, as well as Stata and MATLAB. Compute capsules can be created from existing GitHub folders by easy drag and drop.

Code Ocean staff will assist you in generating a compute capsule for your code and a working copy of this compute capsule will be used in the peer review process, once it is determined that everything runs. If you have selected Double Blind Peer Review, we will make sure the capsule contains no information about your identity.

Should your paper be accepted for publication, Code Ocean will assign a Digital Object Identifier (DOI) to your compute capsule. It will then be embedded into your article. Code Ocean, through [CLOCKSS](https://www.clockss.org/clockss/Home), will guarantee the preservation of all elements of Code Ocean's compute capsules, including the code, data, results, metadata, Dockerfile and Docker image (computational environment) associated with your paper.

By using this platform, other researchers will then be able to easily find and run the code, as well as build upon your work, without any additional setup or configuration of the software. It will also enable preservation of your code, data and the complete environment so that the code associated with this publication is maintained. Should the paper be rejected, you will retain full control over the compute

capsule, and be able to decide what to do with it (publish it, modify it etc).

Please let us know if you would like to set up a compute capsule and we will let you know about the next steps.

Alternatively, we ask you to complete the Software and custom code submission checklist below and return it, and any necessary materials:

<https://www.nature.com/authors/policies/Software.pdf>

Please note that this form is a dynamic 'smart pdf' and must therefore be downloaded and completed in Adobe Reader. Also, we request that you deposit code in a DOI-minting repository such as Zenodo, Gigantum or Code Ocean and cite the DOI in the Reference list. We also request that you use code versioning and provide a license.

For more information on our code sharing policy and requirements, please see:

<https://www.nature.com/nature-research/editorial-policies/reporting-standards#availability-of-computer-code>

OPEN SCIENCE REQUIREMENTS

REPORTING SUMMARY AND EDITORIAL POLICY CHECKLISTS

When revising your manuscript, please update your reporting summary and editorial policy checklists.

Reporting summary: <https://www.nature.com/documents/nr-reporting-summary.zip>

Editorial policy checklist: <https://www.nature.com/documents/nr-editorial-policy-checklist.zip>

If your paper includes custom software, we also ask you to complete a supplemental reporting summary.

Software supplement: <https://www.nature.com/documents/nr-software-policy.pdf>

Please submit these with your revised manuscript. They will be available to reviewers to aid in their evaluation if the paper is re-reviewed. If you have any questions about the checklist, please see <http://www.nature.com/authors/policies/availability.html> or contact me.

Please note that these forms are dynamic 'smart pdfs' and must therefore be downloaded and completed in Adobe Reader. We will then flatten them for ease of use by the reviewers. If you would like to reference the guidance text as you complete the template, please access these flattened versions at <http://www.nature.com/authors/policies/availability.html>.

Finally, please ensure that you retain unprocessed data and metadata files after publication, ideally archiving data in perpetuity, as these may be requested during the peer review and production process or after publication if any issues arise.

DATA AVAILABILITY

We strongly encourage you to deposit all new data associated with the paper in a persistent repository where they can be freely and enduringly accessed. We recommend submitting the data to discipline-specific and community-recognized repositories; a list of repositories is provided here: <http://www.nature.com/sdata/policies/repositories>

All novel DNA and RNA sequencing data, protein sequences, genetic polymorphisms, linked genotype and phenotype data, gene expression data, macromolecular structures, and proteomics data must be deposited in a publicly accessible database, and accession codes and associated hyperlinks must be provided in the "Data Availability" section.

Refer to our data policies here: <https://www.nature.com/nature-research/editorial-policies/reporting-standards#availability-of-data>

To further increase transparency, we encourage you to provide, in tabular form, the data underlying the graphical representations used in your figures. This is in addition to our data-deposition policy for specific types of experiments and large datasets. For readers, the source data will be made accessible directly from the figure legend. Spreadsheets can be submitted in .xls, .xlsx or .csv formats. Only one (1) file per figure is permitted: thus if there is a multi-paneled figure the source data for each panel should be clearly labeled in the csv/Excel file; alternately the data for a figure can be included in multiple, clearly labeled sheets in an Excel file. File sizes of up to 30 MB are permitted. When submitting source data files with your manuscript please select the Source Data file type and use the Title field in the File Description tab to indicate which figure the source data pertains to.

Please include a "Data availability" subsection in the Online Methods. This section should inform readers about the availability of the data used to support the conclusions of your study, including accession codes to public repositories, references to source data that may be published alongside the paper, unique identifiers such as URLs to data repository entries, or data set DOIs, and any other statement about data availability. At a minimum, you should include the following statement: "The data that support the findings of this study are available from the corresponding author upon request", describing which data is available upon request and mentioning any restrictions on availability. If DOIs are provided, please include these in the Reference list (authors, title, publisher (repository name), identifier, year). For more guidance on how to write this section please see: <http://www.nature.com/authors/policies/data/data-availability-statements-data-citations.pdf>

MATERIALS AVAILABILITY

As a condition of publication in Nature Methods, authors are required to make unique materials promptly available to others without undue qualifications.

Authors reporting new chemical compounds must provide chemical structure, synthesis and characterization details. Authors reporting mutant strains and cell lines are strongly encouraged to use established public repositories.

More details about our materials availability policy can be found at <https://www.nature.com/nature-portfolio/editorial-policies/reporting-standards#availability-of-materials>

Please do not hesitate to contact me if you have any questions or would like to discuss these revisions

further. We look forward to seeing the revised manuscript and thank you for the opportunity to consider your work.

Sincerely,

Hui Hua, Ph.D.
Associate Editor
Nature Methods

Reviewers' Comments:

Reviewer #1:

Remarks to the Author:

In this manuscript, the authors present a new method for host depletion that outperforms existing published and commercial methods convincingly, especially on solid tissue. This solves an important problem in the field and will be of broad utility to a wide audience.

A couple of minor points:

- Readers will want to know what are the specific taxa that drop out in Fig. 3E and if there are any commonalities that explain these dropouts e.g. in membrane composition.
- It is puzzling that the QIAGEN viral kit not the microbiome kit was used, when the results focus on bacteria. At least some explanation of why this was done should be provided.
- Per Fig. 1, the protocol does not actually work as well as lyPMA for saliva (except when DTT is added) or as well as QIAamp for soft tissue, at least in terms of host depletion (sequencing results are not provided for these comparisons as far as I can tell). The claims should therefore be refined to reflect this reality, as readers are likely to be left with the inaccurate impression that the new method works better for all sample types. This does not detract from the impressive results in hard tissue.

Overall, this is an important paper and is likely to be broadly appreciated.

Reviewer #2:

Remarks to the Author:

This study by Dr Ismagilov and co-authors developed a comprehensive microbial-enrichment method (MEM) to remove host nucleic acids from complex samples. The performance of MEM was compared to a number of current approaches for different sample types, and this method could be applied to human GI biopsies to characterize microbial taxa, metabolic modules, and construct metagenome-assembled genomes. Overall, the study is instructive and well-organized.

Some comments are as follow:

1. This article sometimes overstates MEM. For example, "MEM is the first method to enable over 1,000-fold removal of host DNA from solid mammalian tissue while minimally perturbing the microbial

community composition". To my knowledge, although there is no commercially available kit for direct tissue depletion, there are some studies reporting the methods of host DNA depletion (Microbiologyopen. 2021 Jun;10(3):e1191. Genomics Proteomics Bioinformatics. 2022 Sep 26:S1672-0229(22)00119-X). The authors should avoid overstatements and address these published methods in Discussion.

2. In addition, commercial kits can achieve good DNA extraction efficiency in tissue samples with some modifications, such as trypsin digestion prior to QIAamp kit (Science. 2016 Nov 11;354(6313):744-747). This should also be discussed.

3. It is impressive that MEM has superior performance when compared with two commercial kits and lyPMA. I think the authors should also consider comparing MEM to a highly cited method using 2.2% saponin, which allows up to 99.99% of host nucleic acid removal in clinical samples (Nat Biotechnol. 2019 Jul;37(7):783-792). Although saponin is present in AHL buffer of QIAamp, its concentration (7.73 wt-%) is much higher.

4. The authors stated that "We optimized every aspect of MEM (bead-beating, incubation step, enzymatic NA removal, enzyme inactivation, handling, etc.) to keep the entire protocol time under 20 min, with gentle processing conditions to prevent microbe lysis. (Methods; Figure S1)." I think Figure S1 could not fully demonstrate the above factors. What are the details of enzyme inactivation and handling? Please also clarify why DNase was used in Figure S1A but not Benzonase.

5. The MoLYsis+PK group in Figure 1C seems to be unnecessary. Although PK is recommended for solid samples, homogenized stool samples should not be considered as solid.

6. "We have analyzed biopsies with as few of 10216S rRNA gene copies/ μ L". Does the " μ L" refer to the volume of nuclease-free water that samples were eluted into? Perhaps it would be clearer to use 16S copies/mg tissue or bacteria number/mg tissue.

7. Please provide the detailed list of 90% "unchanged" taxa.

8. On Page 13, please change any redundancy to "contamination" to prevent misunderstandings.

9. The authors should elaborate how the gene detection score was calculated.

10. Although this article provides a method of benchmark depletion efficiency by faecal sample, a mock community may perhaps a better technique (e.g., use different ratios of bacteria by bacteria free host cell line or microbiome-free tissue), which may be worth for discussion.

11. Contaminant DNA has a negative impact on microbiome research, particularly when evaluating samples with low microbial biomass. It can come from a variety of sources, including laboratory surroundings, researchers, and laboratory reagents. The authors used at least three processing blanks for each DNA extraction batch as a baseline for host DNA detection, but did not set blank controls at the host-DNA depletion step to monitor microbial contamination from the 1.4mm beads-beating procedure. To ensure that the increased taxon (compared to non-host-depletion groups) identified in MEM-treated samples is not the result of contamination, blank controls for MEM should be established.

Reviewer #3:

Remarks to the Author:

Here, we developed and optimized a microbial-enrichment method (MEM) to remove host nucleic acids from complex samples while not substantially perturbing the microbial community composition. We demonstrate performance of MEM in laboratory and clinical settings, and with a range of sample types, including saliva, feces, intestinal scrapings, and intestinal biopsies. We demonstrate the ability of MEM followed by shotgun metagenomic sequencing to detect both high- and low-abundance microbial taxa, metabolic modules (defined as sub-pathways as annotated by KEGG24), and genes from human intestinal biopsies along the GI tract.

- But I am missing the results of negative/blank controls. Particularly, if you want to include the low abundant taxa into your output. What is the kitome of your workflow? [1, 2, 3, 4].

[1] Weyrich LS, Farrer AG, Eisenhofer R, Arriola LA, Young J, Selway CA, Handsley-Davis M, Adler CJ, Breen J, Cooper A. Laboratory contamination over time during low-biomass sample analysis. *Mol Ecol Resour.* 2019 Jul;19(4):982-996. doi: 10.1111/1755-0998.13011. Epub 2019 Apr 29. PMID: 30887686; PMCID: PMC6850301.

[2] Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, Turner P, Parkhill J, Loman NJ, Walker AW. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol.* 2014 Nov 12;12:87. doi: 10.1186/s12915-014-0087-z. PMID: 25387460; PMCID: PMC4228153.

[3] Velásquez-Mejía EP, de la Cuesta-Zuluaga J, Escobar JS. Impact of DNA extraction, sample dilution, and reagent contamination on 16S rRNA gene sequencing of human feces. *Appl Microbiol Biotechnol.* 2018 Jan;102(1):403-411. doi: 10.1007/s00253-017-8583-z. Epub 2017 Oct 27. PMID: 29079861.

[4] Liu Y, Elworth RAL, Jochum MD, Aagaard KM, Treangen TJ. De novo identification of microbial contaminants in low microbial biomass microbiomes with Squeegie. *Nat Commun.* 2022 Nov 10;13(1):6799. doi: 10.1038/s41467-022-34409-z. PMID: 36357382; PMCID: PMC9649624.

- I am also missing positive controls with a known mixture and known concentration of Gram-positive and Gram-negatives, acid-fast bacteria, fungi etc. What is the effect of your proposed protocol on a broader range of community members?
- Although the results are positive (Figure 1D), the authors do not explain/discuss why MEM performs better than other methods. For instance, when using detergents to selectively lyse host cells, some bacterial cells may also be lysed due to differences in cell wall structures among bacterial species. In contrast, the big bead beating method selectively lyses cells much bigger than bacterial cells, reducing the level of bias between different bacterial species. For each host lysis method, the authors should also determine if the differences between treatment/control are linked to taxonomy (e.g., replot Figure 3D as several sub-histogram, each for a different phylum).
- The estimation of bacterial DNA loss requires further clarification. The value displayed in Figure 1C does not directly reflect bacterial DNA loss, but rather represents the overall output of the host-DNA depletion methods and the subsequent bacterial DNA extraction process. The current conclusions seem to presume that the downstream bacterial DNA extraction method effectively retrieves all DNA molecules in a sample, which is not a valid assumption. Additionally, certain pre-treatment steps during the host-DNA depletion process may even enhance the efficiency of downstream bacterial DNA extraction (Figure 2S).
- "About half (91 of 187) of the microbial species identified were unique to an individual (Fig 4d)." This is a potentially significant finding, as species from stool are not this different between people, and the authors should elaborate further on this observation. What are the specific species that are unique to each individual? What are the species that are present in multiple individuals? Is this result merely

due to the limited number of bacterial cells present in each biopsy sample?

- The outcomes from the metabolic modules are greatly impacted by the current database and methodology, thus it's hard to assess the significance of the results. Additionally, for gut species, usually ~50% of genes can typically be annotated. Therefore, the fact that the overwhelming majority of genes from MEM metagenomes are unannotated raises concerns about the accuracy of the computational analysis (Figure 4A).
- The authors assess the quality of MAGs using completeness and contamination, which is insufficient in many ways. How does the MAG size compare to the typical genome size of isolates from that species? Are the MAGs containing 16S, 23S, and 5S sequences? (Also check the MIMAG reference: <https://doi.org/10.1038/nbt.3893>.)
- Detection of SNVs (SNPs) by comparing MAGs is not a reliable method as the assembly process often introduces numerous errors. The conventional way to detect SNVs is to use sequence alignment tools such as bowtie2 and samtools. The authors should refer to the many papers calling SNVs from metagenomic datasets (e.g., doi.org/10.1371/journal.pbio.3000102). The use of the fixation index requires certain assumptions to be met, and it is unclear if those assumptions are satisfied in this case. The authors should either clarify the concept of a fixation index and why it is appropriate to use here, or opt for a simpler metric, such as counting the number of SNVs between samples, which could still lead to the same conclusion.
- The method used for normalizing host depletion is not described (e.g., see Figures 1E, 1F, 1G). According to the method section, each sample may have different starting weights or volumes, resulting in varying amounts of host DNA. Normalization criteria may also differ based on the type of sample. It is unclear whether the number of genomes remaining is based on a specific volume of DNA extract, the entire sample, or some other definition.
- The use of detection score might raise the concern that the MAGs may include contigs that are contaminations and detection score cannot tell if that's the case. The authors should use a metric like absolute of normalized sequencing depth or provide more justification for using detection score.

For saliva, we observed a high correlation in the relative abundances of microbial taxa between the control sample and both the MEM and MEM+DTT treatments (Fig 2d). The correlation was less pronounced for low-abundance taxa with enrichment of specific species in MEM-treated samples

- I would like to see the negative controls here. What is the background noise of MEM treatment? Is the observed enrichment of specific low abundance species signal or noise?

MEM-treated saliva samples had lower CV (50% CV, 95% CI), indicating better replicability compared with untreated controls.

- Again, if there is a persisting contamination background signal of certain taxa in MEM-treated samples, the replicability increases because the background signal has a high recognition value.

we shotgun-sequenced paired control and MEM-treated biopsies from CT18 at a depth of above 100 million reads (N=2 for each condition) (Fig 3a). We observed roughly a 100-fold increase in the number of organisms that could be detected

- Again, I would like to see the blank, negative controls here for MEM-treated biopsies and controls. Detecting more taxa is not necessarily beneficial, if it's just more contamination.

Figure 4: Shotgun sequencing of MEM-treated human intestinal biopsies

- Please show and quantify whether individual patterns persist as a function of time instead of showing all the individual PCA plots. Why is there so much variance in some individuals (Figure 4), e.g., for CT13, figure 4G, one descending colon sample is found on the complete opposite site of the x-axis compared to the other two descending colon samples with a variance of about 60%. Also, I only detect two descending colon samples for CT13 (figure 4F). Where is the third one? For CT14, 4F I only detect 1 rectum sample (color red). Where are the others? Why is there sometimes more variance (PC1) on highly compressed module than gene level (CT8, CT12, CT13)?

Data Availability

The datasets generated during and analyzed during the current study will be made available upon publication at CaltechDATA, URL pending. Microbial sequencing data will be made available upon publication at NCBI Accession doi pending.

Code Availability

The code utilized in data processing and analysis will be made available upon publication at CaltechDATA, URL pending.

- It would be great to share the code for the review process as well.

Author Rebuttal, first revision:

Point-By-Point Response to Reviewers

(New additions to the manuscript and SI are indicated in red text)

Reviewer #1:

Remarks to the Author:

In this manuscript, the authors present a new method for host depletion that outperforms existing published and commercial methods convincingly, especially on solid tissue. This solves an important problem in the field and will be of broad utility to a wide audience.

A couple of minor points:

-Readers will want to know what are the specific taxa that drop out in Fig. 3E and if there are any commonalities that explain these dropouts e.g. in membrane composition.

We have added supplemental files (Table S1-3, S6) documenting all taxonomic changes between host-depleted samples (including MEM) and control samples for Figs 1C-D/2B (mouse stool), 2D-E (human saliva), and 3C-E (human colonic biopsies) obtained from both 16S rRNA gene and shotgun sequencing. It was difficult to identify any commonalities between taxa dropouts as many of these taxa appear to drop out in some participants' samples but are retained in other participants' samples, potentially indicating that in controls, these taxa were detected from extracellular DNA corresponding to dead bacteria. Readers will be able to search these tables to see how specific taxa of interest are affected by each host depletion method in each sample type.

- It is puzzling that the QIAGEN viral kit not the microbiome kit was used, when the results focus on bacteria. At least some explanation of why this was done should be provided.

The QIAGEN PowerViral extraction kit has been used extensively in microbiome studies (see below citations). Additionally, comparisons of DNA extraction kits have shown QIAGEN PowerViral kit to yield higher levels of microbial DNA (see Mancabelli *et al.* 2022). The optional mechanical lysis step was also performed to ensure successful bacterial lysis.

- Mahajan, *et al.* Vaginal microbiome-host interactions modeled in a human vagina-on-a-chip. *Microbiome* 10, 201 (2022).
- Palmieri, *et al.* Microbiome Analysis of Mucosal Ileoanal Pouch in Ulcerative Colitis Patients Revealed Impairment of the Pouches Immunometabolites. *Cells*. 10(11):3243. doi: 10.3390/cells10113243 (2021).
- Calusinska, *et al.* Integrative omics analysis of the termite gut system adaptation to *Miscanthus* diet identifies lignocellulose degradation enzymes. *Commun Biol* 3, 275 (2020).
- Mancabelli, *et al.* Guideline for the analysis of the microbial communities of the human upper airways. *Journal of Oral Microbiology* 14, 2103282 (2022).

The Methods section has been updated to add this clarification and the Mancabelli citation:

Nucleic acids were isolated following Qiagen's AllPrep PowerViral DNA/RNA Kit (Cat # 28000-50). Samples were homogenized in 0.1 mm glass beads for 1 min at 6m/s using FastPrep-24 (MP Biomedical) to ensure complete microbial lysis.⁵⁷

57. Mancabelli, L. et al. Guideline for the analysis of the microbial communities of the human upper airways. *Journal of Oral Microbiology* 14, 2103282 (2022).

- Per Fig. 1, the protocol does not actually work as well as lyPMA for saliva (except when DTT is added) or as well as QIAamp for soft tissue, at least in terms of host depletion (sequencing results are not provided for these comparisons as far as I can tell). The claims should therefore be refined to reflect this reality, as readers are likely to be left with the inaccurate impression that the new method works better for all sample types. This does not detract from the impressive results in hard tissue.

We have updated the language appropriately below:

In saliva, all methodologies enabled some host removal. Following MEM treatment, over 40-fold depletion of host was achieved (Fig 1E). The addition of DTT pre-treatment, which was added due to the high mucin content of saliva, slightly increased host removal by MEM in some participants (Fig 1E, Fig S3). lyPMA appeared similarly effective at host removal as MEM with the DTT pretreatment, but was difficult to use predictably because the stoichiometric nature of the method can result in large microbial losses when host levels are lower than expected (Fig S3).

We next examined host depletion on whole tissue samples, beginning with mouse intestinal scrapings that isolate the epithelial layer with mucosa-associated bacteria (Fig S4). Mucosal scraping samples were efficiently host-depleted by MEM and some of the published methods (Fig 1F). MEM, MoLYsis, and QIAamp all showed around 1000-fold depletion of host with QIAamp showing slightly greater host-removal (MEM had an average 1,600-fold depletion [SD: 170]). lyPMA performed poorly on the soft tissue sample because this method relies on UV-activated crosslinking making it incompatible with opaque sample types.

Overall, this is an important paper and is likely to be broadly appreciated.

We appreciate the Reviewer's recognition of the broad utility of this new method.

Reviewer #2:

Remarks to the Author:

This study by Dr Ismagilov and co-authors developed a comprehensive microbial-enrichment method (MEM) to remove host nucleic acids from complex samples. The performance of MEM was compared to a number of current approaches for different sample types, and this method could be applied to human GI biopsies to characterize microbial taxa, metabolic modules, and construct metagenome-assembled genomes. Overall, the study is instructive and well-organized.

Some comments are as follow:

1. This article sometimes overstates MEM. For example, "MEM is the first method to enable over 1,000-fold removal of host DNA from solid mammalian tissue while minimally perturbing the microbial community composition". To my knowledge, although there is no commercially available kit for direct tissue depletion, there are some studies reporting the methods of host DNA depletion (Microbiologyopen. 2021 Jun;10(3):e1191. Genomics Proteomics Bioinformatics. 2022 Sep 26:S1672-0229(22)00119-X). The authors should avoid overstatements and address these published methods in Discussion.

We thank the Reviewer for sharing this excellent paper by Bruggeling *et al.* (Microbiologyopen. 2021 Jun;10(3):e1191.) However, when looking through their results, host DNA was removed 1.9-fold in human tissues (from Figure 3c). Additionally, when looking at their SI, they report a maximum of 135-fold removal of host DNA in fish samples (Table A3). We therefore do not think our claim is overstated. However, we have added this citation into our discussion of existing host-removal methodologies. We also note the potential value of incorporating enzymatic pre-treatments (such as mutanolysin as used in Bruggeling *et al.*) to assist in bacterial recovery and, to avoid arguments about these details, we rephrased the sentence in question and also emphasized the ability to construct MAGs.

We have made the following changes in the Introduction:

To prevent the majority of shotgun-sequencing reads assigning to host, a wide variety of host-removal (i.e. host- depletion) methods have been developed²¹⁻²⁷; however, commercial and published protocols are primarily used on liquid samples, and none have been shown to be sufficiently effective **to enable bacterial MAG construction from** solid mammalian tissues.

We have made the following changes in the Results:

lyPMA appeared similarly effective at host removal as MEM but was difficult to use predictably because the stoichiometric nature of the method can result in large microbial losses when host levels are lower than expected (Fig S2). **Additionally, MoLYsis showed increased bacterial recovery, likely due to the additional mutanolysis step.**^{22,26}

22. Bruggeling, C. E. et al. Optimized bacterial DNA isolation method for microbiome analysis of human tissues. *Microbiologyopen* 10, e1191, doi:10.1002/mbo3.1191 (2021).

26. Cheng, W. Y. et al. High Sensitivity of Shotgun Metagenomic Sequencing in Colon Tissue Biopsy by Host DNA Depletion. *Genomics, proteomics & bioinformatics*, doi:10.1016/j.gpb.2022.09.003 (2022).

We have made the following changes in the Discussion:

MEM ~~is the first method to~~ enables over 1,000-fold removal of host DNA from solid mammalian tissue while minimally perturbing the microbial community composition.

2. In addition, commercial kits can achieve good DNA extraction efficiency in tissue samples with some

modifications, such as trypsin digestion prior to QIAamp kit (Science. 2016 Nov 11;354(6313):744-747). This should also be discussed.

We appreciate the Reviewer bringing up this excellent paper, which we now cite. We were unable to find any data quantifying host removal from this paper by Avanzi et al. (Science. 2016 Nov 11;354(6313):744-747). The authors state they performed qPCR of the squirrel mitochondrial DNA, but the data was not provided. Therefore, it is difficult to determine how effective trypsin digestion was. Furthermore, this paper does not report MAG construction.

However, the Bruggeling *et al.* paper (mentioned in the prior comment) involved Proteinase K treatment followed by saponin-selective lysis of host. Because the QIAamp kit is also a saponin-based lysis method, we would expect the selective lysis of host to be similarly effective. For the protease pretreatment, we expect Proteinase K to be similar if not more effective than a trypsin digestion because Proteinase K is known to produce shorter peptides than trypsin (see Dau et al. 2020; PMC7377536). Given that the Proteinase K pretreatment followed by saponin-selective lysis did not produce comparable host removal to MEM, we would expect the same to be true of trypsin digestion prior to QIAamp. However, we chose not to get into these speculative discussions in the revised manuscript. We now cite this paper (Avanzi et al. 2016) and emphasize MAG construction by MEM.

We have added citations to both papers in the manuscript:

To prevent the majority of shotgun-sequencing reads assigning to host, a wide variety of host-removal (i.e. host- depletion) methods have been developed²¹⁻²⁶; however, commercial and published protocols are primarily used on liquid samples, and none have been shown to be sufficiently effective to enable bacterial MAG construction from solid mammalian tissues.

22. Bruggeling, C. E. et al. Optimized bacterial DNA isolation method for microbiome analysis of human tissues. *Microbiologyopen* 10, e1191, doi:10.1002/mbo3.1191 (2021).

24. Avanzi, C. et al. Red squirrels in the British Isles are infected with leprosy bacilli. *Science* 354, 744-747, doi:10.1126/science.aah3783 (2016).

25. Charalampous, T. et al. Nanopore metagenomics enables rapid clinical diagnosis of bacterial lower respiratory infection. *Nature Biotechnology* 37, 783-792, doi:10.1038/s41587-019-0156-5 (2019).

26. Cheng, W. Y. et al. High Sensitivity of Shotgun Metagenomic Sequencing in Colon Tissue Biopsy by Host DNA Depletion. *Genomics, proteomics & bioinformatics*, doi:10.1016/j.gpb.2022.09.003 (2022).

3. It is impressive that MEM has superior performance when compared with two commercial kits and lyPMA. I think the authors should also consider comparing MEM to a highly cited method using 2.2% saponin, which allows up to 99.99% of host nucleic acid removal in clinical samples (*Nat Biotechnol.* 2019 Jul;37(7):783-792). Although saponin is present in AHL buffer of QIAamp, its concentration (7.73 wt-%) is much higher.

We thank the Reviewer for their recommendation and providing the reference to Charalampous *et al.* (*Nat Biotechnol.* 2019 Jul;37(7):783-792). We would expect a higher concentration of saponin to be more effective at lysing host cells at the expense of inducing further microbial bias. Because we have tested the higher concentration (7.73 wt%) of saponin already (through the QIAamp kit), we would not predict a lower concentration (2.2 wt%) to improve host removal, although it may reduce bacterial losses induced by QIAamp.

We have updated a statement in the Results as follows:

By comparing paired host-depleted and control samples, we found that lyPMA and QIAamp induced the most total bacterial losses whereas MoLYsis and QIAamp induced the least uniform bacterial losses, with some taxa dropping more than 100-fold (Fig 1C-D). Previous literature suggests QIAamp's saponin concentration can be lowered to limit some of these bacterial losses.²⁵

25. Charalampous T, et al. Nanopore metagenomics enables rapid clinical diagnosis of bacterial lower respiratory infection. *Nat Biotechnol.* 2019

Jul;37(7):783-792.

We would also like to point out that for mucosal scrapings, our QIAamp results were very similar to the Charalampous *et al.* results, demonstrating over 99.98% removal of host DNA. Because Charalampous *et al.* used respiratory sputum, secretions, and aspirates, these sample compositions are likely more akin to mucosal scrapings rather than tissue biopsies.

4. The authors stated that “We optimized every aspect of MEM (bead-beating, incubation step, enzymatic NA removal, enzyme inactivation, handling, etc.) to keep the entire protocol time under 20 min, with gentle processing conditions to prevent microbe lysis. (Methods; Figure S1).” I think Figure S1 could not fully demonstrate the above factors. What are the details of enzyme inactivation and handling? Please also clarify why DNase was used in Figure S1A but not Benzonase.

To clarify our enzyme inactivation and handling procedures we have made the following edits to the Methods:

Microbial-Enrichment Method (MEM). Samples for MEM treatment were placed into 2-mL 1.4-mm ceramic bead- beating tubes (Lysing Matrix D from MP Biomedical, Cat #116913050-CF) with a maximum volume of 400 μ L. For solid sample types (stool and intestinal tissue), up to 400 μ L of saline (0.9% NaCl, autoclaved) was added into the bead-beating tube. Samples were homogenized using FastPrep-24 for 30 sec at 4.5 m/sec and then immediately placed on ice. A total of 150 μ L of homogenized tissue was removed and placed into a clean microfuge tube containing 10 μ L of buffer (100 mM Tris + 40 mM $MgCl_2$, pH 8.0 and 0.22 μ m sterile filtered), 33 μ L of saline (0.9% NaCl, autoclaved), 2 μ L of EMD Millipore Benzonase Nuclease HC (Cat # 71205), and 5 μ L of NEB Proteinase K (Cat # P8107S). Samples were mixed lightly by manually pipetting up and down 5-10 times and spun briefly to pool (1,000 x g for 5 seconds). Tubes were placed on a dry block incubator for 15 min at 37 °C with shaking at 600 rpm. Samples were then pelleted at 10,000 xg for 2 min and the supernatant was removed and discarded. Pellets were resuspended in 150 μ L of PrimeStore MTM (Longhorn), **a transport medium, to inactivate residual enzymatic activity** and stored at -80 °C until nucleic-acid extraction.

We appreciate Reviewer’s careful reding of the paper. We did not use DNase in this paper except in early experiments (in Fig. S1A); the data in the main paper are all with Benzonase. We switched between

DNase and Benzonase early in the protocol optimization process out of concern regarding enzyme inactivation. During optimization, we observed that some microbial lysis continued during DNase heat inactivation so we switched to using Benzonase, which we found is fully inactivated by PrimeStore MTM. We have seen DNase and Benzonase to be similarly effective at removal of nucleic acids. We have updated the language in the methods to reflect this lesson learned:

The initial MEM protocol utilized DNase I treatment in place of Benzonase. However, we noted continuous microbial lysis during DNase I heat inactivation. Benzonase was used to remove high heat steps as it is fully inactivated by PrimeStore MTM.

We have made the following changes to the Results:

We optimized many aspects of MEM, including enzymatic NA removal, bead-beating, and incubation time, to keep the entire protocol time under 20 min, with gentle processing conditions to prevent microbe lysis. (Methods; Fig S1).

5. The MoLYsis+PK group in Figure 1C seems to be unnecessary. Although PK is recommended for solid samples, homogenized stool samples should not be considered as solid.

We agree that MoLYsis+PK group is not absolutely critical for illustrating analysis of stool samples. We included it for a different reason. It illustrates the effect of PK on microbial losses and stool is a reasonable sample to use here for two reasons: stool samples are more homogenous than colonic sections, and stool samples contain many more taxa than mucosal scrapings. We would prefer to leave this condition for continuity with Figure 1F+1G, however if the Reviewer and Editor feel strongly about taking it out, we'd take it out.

6. "We have analyzed biopsies with as few of 10216S rRNA gene copies/ μ L". Does the " μ L" refer to the volume of nuclease-free water that samples were eluted into? Perhaps it would be clearer to use 16S copies/mg tissue or bacteria number/mg tissue.

We have clarified what this number refers to and added the number of 16S copies per mg of tissue for additional clarity.

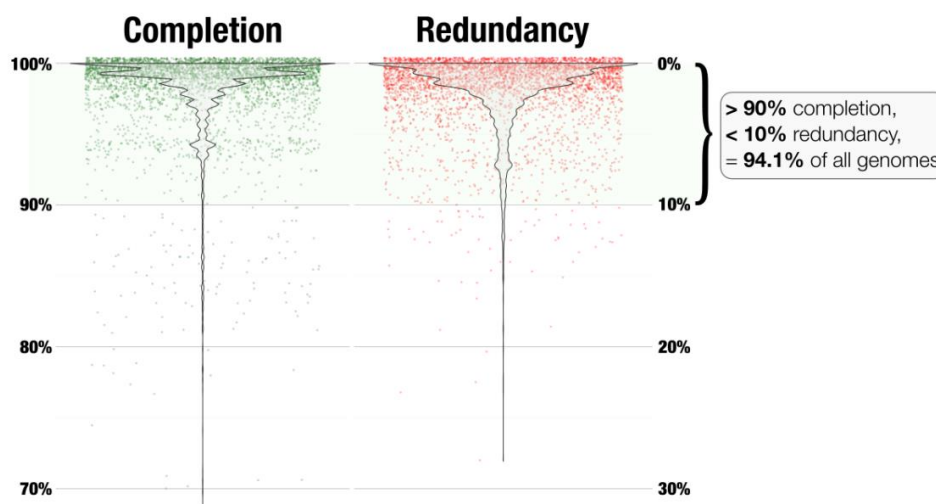
We acknowledge the following limitations of the MEM approach demonstrated here. We have analyzed biopsies with as few of 10^2 16S rRNA gene copies/ μL in the 100 μL elution from the extraction column (corresponding to $\sim 10^4$ 16S copies/mg of tissue), however deep analysis of samples below this bacterial load will require greater levels of host depletion and/or greater sequencing depth (Fig S6-7).

7. Please provide the detailed list of 90% “unchanged” taxa.

We have added supplemental Tables S1-S3, S6 that shows the log2 fold-change between these host depleted and control samples. The supplemental tables include both 16S rRNA gene sequencing data and shotgun sequencing data for mouse stool, human saliva, and human intestinal biopsies.

8. On Page 13, please change any redundancy to “contamination” to prevent misunderstandings

We thank the Reviewer for bringing up this point and appreciate the opportunity to explain our rationale for using the term “redundancy” rather than “contamination.” The word “Redundancy” essentially aims to communicate the multiple occurrences of single-copy core genes (SCGs). Single-copy core gene collections are generated by surveying all available genomes and by identifying genes that occur in a large fraction of genomes (i.e., >95%) as a single-copy. However, not every genome has all of these core genes, and some genomes will have these genes as multi-copy. For instance, here is an analysis of over 4,000 circular bacterial genomes from the NCBI, using a single copy-core gene collection published by Campbell et al. (<https://www.pnas.org/doi/full/10.1073/pnas.1303090110>):



This figure shows that many of the complete and circular genomes, which represent the highest quality of genomes with no contamination, can have many of the ‘single-copy core genes’ in multiple copies. Which means, while the multiple occurrences of SCGs may be a good indicator of contamination, it is an assumption that one can’t be certain of until it is tested. Thus, for broad scale analyses of quality of MAGs through SCGs, the software platform we used in our analyses, anvi’o (<https://www.nature.com/articles/s41564-020-00834-3>), reports completion and redundancy estimates rather than completion and contamination estimates for MAGs. Thus, to prevent misleading conclusions regarding actual contamination due to the occurrence of multi-copy SCGs, we prefer to continue using the term redundancy.

9. The authors should elaborate how the gene detection score was calculated.

This is a good catch, and we apologize for the confusion in the initial manuscript submission. The term “detection score” was incorrectly used in place of the term “detection.” We have fixed this error throughout the manuscript to promote continuity with previously published analyses. We have also updated the Methods section as follows with further details on how detection is calculated:

Pangenome Analysis. A *Phocaeicola vulgatus* MAG from the terminal ileum of CT12 was selected as a reference genome based on genome length. Open reading frames (ORFs) were identified through Prodigal for the *P. vulgatus* reference genome. Non-host reads from each participant (CT7, CT12, CT13, and CT14) were mapped onto the *P. vulgatus* reference genome by following anvi’o’s metagenomics workflow using reference mode. For each sample and each gene

present in the *P. vulgatus* reference genome, gene detection was calculated. Gene detection refers to the percentage of each gene sequence with at least 1X coverage. The average detection across all genes present within the *P. vulgatus* MAG was calculated and samples with a mean detection below 0.25 were removed from the final analysis. Pangenome visualization was performed in anvi'o interactive interface using the gene-mode flag with sorting of samples and genes by detection.

10. Although this article provides a method of benchmark depletion efficiency by faecal sample, a mock community may perhaps a better technique (e.g., use different ratios of bacteria by bacteria free host cell line or microbiome-free tissue),

We understand the appeal of a contrived mock community in method validation. However, even mock communities require comparison to an extracted control due to variation in microbial recovery based on the extraction kit used (see below citations). Therefore, we opted to validate on stool, which includes a larger number of unique taxa and a continuum of relative abundances compared with a mock community. We have added a brief discussion of these points to the Results section:

To quantify how MEM affects microbial community composition and relative abundances of individual taxa, we first used frozen mouse fecal samples. For validation, we chose fecal samples instead of a contrived community to characterize microbial impacts on a wider range of unique taxa and on a continuum of abundances. Additionally, contrived communities still require an extracted control due to variation in extraction kit efficiency^{30,31}. Mouse fecal samples do not typically require host-depletion because they have low levels of host contamination (more than 90% of the DNA biomass originates from non-host cells). Thus, the high biomass of microbial cells in such samples makes them ideal for characterizing the impact of different host-depletion methods on the microbial community composition.

30. Shaffer JP, Carpenter CS, Martino C, Salido RA, Minich JJ, Bryant M, Sanders K, Schwartz T, Humphrey G, Swafford AD, Knight R. A comparison of six DNA extraction protocols for 16S, ITS and shotgun metagenomic sequencing of microbial communities. *Biotechniques*. 2022 Jun;73(1):34-46. doi: 10.2144/btn-2022-0032. Epub 2022 Jun 17. PMID: 35713407; PMCID: PMC9361692.

31. Hallmaier-Wacker, L.K., Lueert, S., Roos, C. *et al.* The impact of storage buffer, DNA extraction method, and polymerase on microbial analysis. *Sci Rep* 8, 6292 (2018). <https://doi.org/10.1038/s41598-018-24573-y>

11. Contaminant DNA has a negative impact on microbiome research, particularly when evaluating samples with low microbial biomass. It can come from a variety of sources, including laboratory surroundings, researchers, and laboratory reagents. The authors used at least three processing blanks for each DNA extraction batch as a baseline for host DNA detection, but did not set blank controls at the host-DNA depletion step to monitor microbial contamination from the 1.4mm beads-beating procedure. To ensure that the increased taxon (compared to non-host-depletion groups) identified in MEM-treated samples is not the result of contamination, blank controls for MEM should be established.

We appreciate this excellent point and agree with it strongly. We added to the paper data for sequenced and quantified bacterial communities in processing blanks. These blanks are samples consisting of saline treated in parallel with the biopsies in the clinic. Blanks were analyzed with Quant-Seq (ddPCR and 16S rRNA gene sequencing) to characterize the degree of contaminant DNA. We have also uploaded SI tables (GIVE NAMES HERE) containing the relative and absolute abundances of taxa within these blanks. The blanks had orders of magnitude lower bacterial signal compared to the saliva and stool samples. For the intestinal biopsies, the bacterial load ranged dramatically (see Fig S5) with some samples having similar bacterial load to the processing blanks. Therefore, 16S rRNA gene sequencing data for intestinal biopsies were blank filtered using MEM processing blanks and we have described what this process entails in the Method (see below). We have updated the data for Figure 3 and 4 with this additional filtering step.

We have made the following changes to the Results:

To determine how MEM performs on human intestinal biopsies, we recruited healthy participants undergoing routine colon cancer screenings via colonoscopy. From each of four participants, we obtained eight mucosal biopsies; four biopsies from each participant were assigned to the MEM-treatment group and four were untreated controls (Fig 3A). Due to concerns regarding contaminant DNA³⁴⁻³⁷ in samples with low bacterial loads, we also characterized the background bacterial signal associated with MEM and our processing methods through quantitative 16S rRNA gene sequencing of MEM processing blanks (see Methods and Tables

S4-S5). MEM removed host DNA more than 2,000-fold across all 16 biopsies, with most biopsies having host levels comparable to a processing blank after MEM treatment. (Fig 3B).

To determine how MEM affects the human intestinal microbiome at a community level, we first performed quantitative 16S rRNA gene sequencing.³³ Roughly 93% of genera remained in MEM-treated and control biopsies after computationally removing taxa found at higher absolute abundances in the blanks, giving us confidence most detected taxa were not background contaminants. To further confirm that MEM did not introduce additional contamination, we found strong agreement between taxon abundances in MEM and control biopsies (Table S6). PCA of sequencing results showed that any differences in microbial relative abundances introduced by MEM were less than the differences observed between participants (Fig 3C). Analysis of sequencing results revealed minimal changes in relative abundances of most taxa after MEM treatment, with roughly 88% of taxa having no significant differences in relative abundances from the controls (Mann Whitney U test, two-sided $P=0.05$). For taxa present at greater than 1% relative abundance, more than 95% of taxa had no significant differences between MEM and control samples. The log2-fold difference in taxa between control and MEM-treated samples approximated a normal distribution (Kolmogorov–Smirnov test against normal distribution, statistic=0.074, $P=0.11$) (Fig 3D). Furthermore, there was a linear correlation in relative taxon abundances between the control and MEM-treated samples (Fig 3E). The speed and simplicity of MEM enables over 1,000-fold host removal while introducing minimal biases in microbial relative abundances when used in a clinical setting on human intestinal biopsies.

We have made the following changes to the Methods:

Longitudinal Sampling of the Human Intestinal Tract (Figure 4C)

For longitudinal sampling, a total of five participants were sampled 12 times from 4 different locations during a routine colonoscopy. The 4 locations sampled were the terminal ileum, ascending colon (designated as 10 cm distal to the ileocecal valve), descending colon, and rectum. From a single field of view (5 cm diameter) from each location, 3 biopsies were collected in one passage with 2.8 mm biopsy forceps and stored dry on ice in a microfuge. For participant CT14, only one

rectal sample was obtained. On average, biopsies were 2.5 mg with a minimum size of 0.1 mg and a maximum of 5.9 mg. All biopsies were then processed individually in the laboratory at University of Chicago (see “MEM”) before shipment on dry ice to Caltech for DNA extraction. Time between specimen collection and processing ranged from 10 min to 52 min. Samples were processed individually in the laboratory at University of Chicago (see “MEM”) before shipment on dry ice to Caltech for DNA extraction. Additionally, three microfuge tubes of 400 μ L saline were opened and closed in the laboratory to serve as clinical processing blanks.

Quant-Seq. Microbial characterization and quantification was performed using the quantitative sequencing (“Quant-Seq”) pipeline we have described previously.³³ Due to the low bacterial loads present in intestinal biopsies, Quant-seq was also performed on three MEM processing blanks. If a taxon was detected at a higher absolute abundance in any of the processing blanks compared to the intestinal biopsies, the taxon was removed from downstream analysis.

We have made the following changes to the Figure 3 and 4:

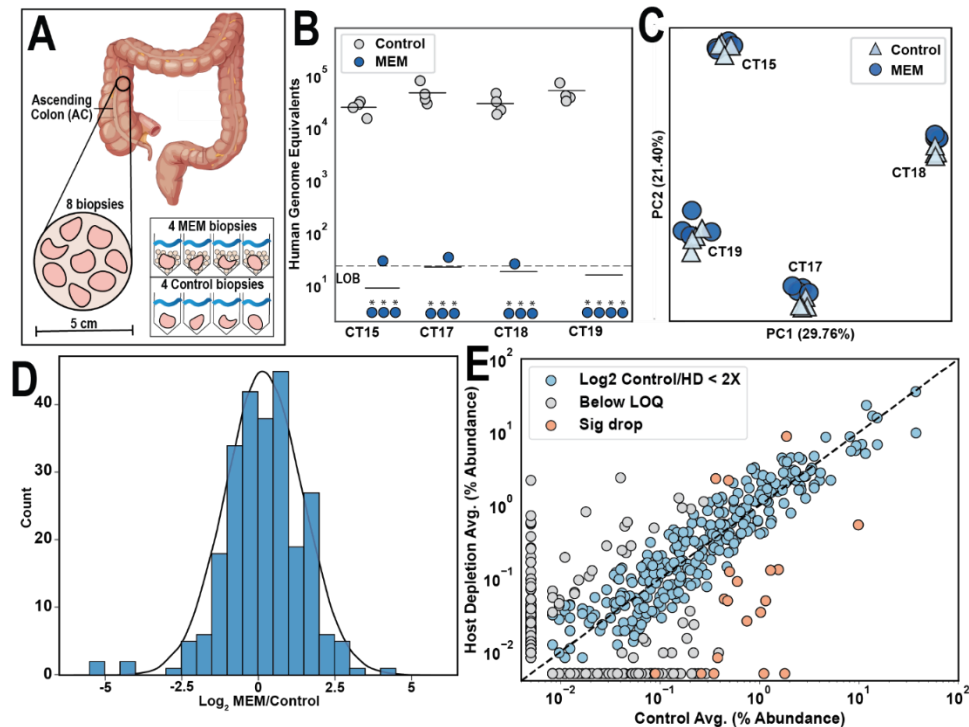


Figure 3: Analysis of microbial enrichment in paired human intestinal biopsies processed with and without MEM A) Sampling graphic illustrates collection from four participants (CT15, CT17, CT18, and CT19) each with eight ascending colon biopsies. Four biopsies from each participant were MEM-treated and four were untreated controls. B) Host DNA was quantified for each biopsy using ddPCR of a single-copy host primer. Human genomes remaining refers to the abundance of this single-copy gene present in 1 μ L of elution (* indicate measurement was below limit of blank [LoB] defined as LoB = meanblank + 1.645[SDblank] based on three processing blanks). C) Biopsies were characterized with 16S rRNA gene sequencing and principal coordinates analysis (PCA) on microbial genus-level relative abundances were performed to visualize microbial population variation. D) Log2-fold differences in relative abundances of taxa between control and MEM-treated biopsies were plotted with a standard normal distribution overlaid in black. E) Relative abundance of taxa measured in control vs MEM-treated biopsies were plotted and overlaid on a dashed line showing 1:1 correlation. Highlighted in gray are taxa that were below the assay limit of quantification (LOQ). Highlighted in orange are taxa with greater than 4-fold changes between control and MEM biopsies.

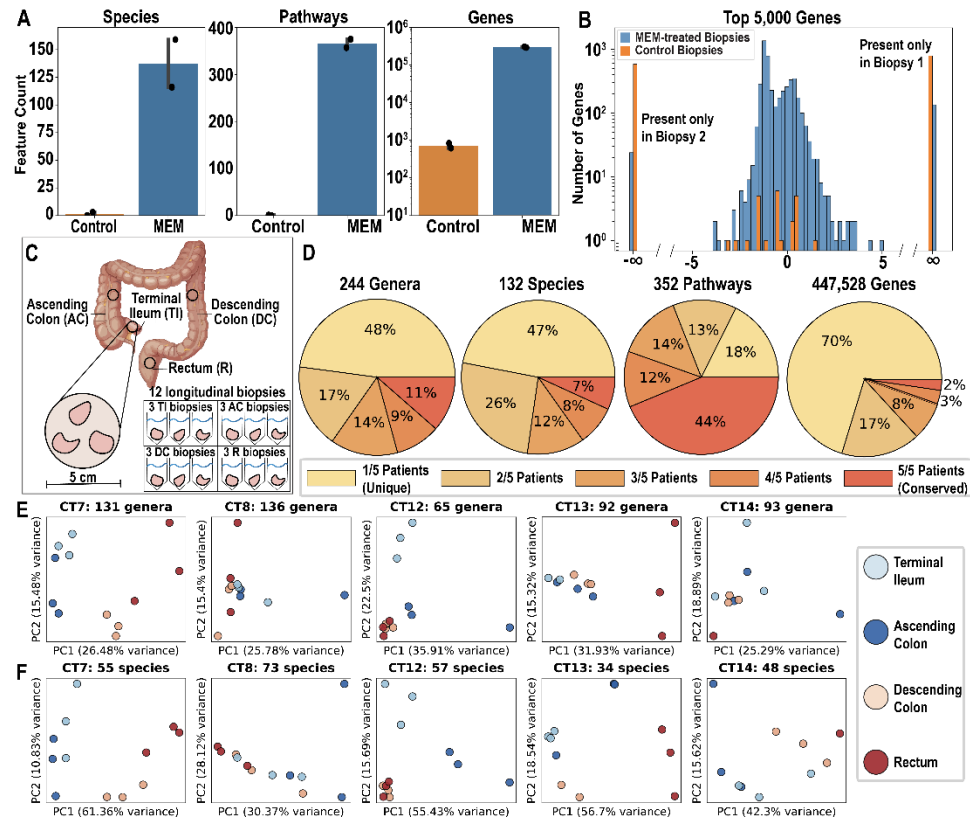


Figure 4: Shotgun sequencing of MEM-treated human intestinal biopsies. A) Four biopsies from participant CT18 (two MEM- treated and two control) were shotgun-sequenced and the number of microbial species, pathways, and genes identified in each sample were plotted (N=2; error bars are 95% CI). B) For the top 5,000 abundant genes, the log₂ fold-change in relative abundances between the two MEM-treated biopsies and the two control biopsies were plotted. C) Sampling graphic for collection of 12 biopsies from each of 5 participants (3 biopsies taken from 4 separate regions of the GI tract). Biopsies within one region were sampled within one field of view (5-cm diameter). D) All 60 biopsies were processed with MEM, followed by 16S rRNA gene sequencing (for genera) and shotgun sequencing (for species, pathways, and genes). The number of features for genera, species, pathways, and genes were grouped based on whether they were present in at least one biopsy sample from only one participant (1/5), two participants (2/5), three participants (3/5), four participants (4/5), or from all participants (5/5). Analysis was performed on annotated genes, and separately on all genes (annotated + unannotated). E-F) PCA was performed on all 60 longitudinal samples grouped by participant (CT7, CT8, CT12, CT13, CT14). E) PCA on relative abundance of 16S rRNA gene sequencing genera assignments. F) PCA on relative abundance of shotgun-sequencing species assignments.

Reviewer #3:**Remarks to the Author:**

Here, we developed and optimized a microbial-enrichment method (MEM) to remove host nucleic acids from complex samples while not substantially perturbing the microbial community composition. We demonstrate performance of MEM in laboratory and clinical settings, and with a range of sample types, including saliva, feces, intestinal scrapings, and intestinal biopsies. We demonstrate the ability of MEM followed by shotgun metagenomic sequencing to detect both high- and low-abundance microbial taxa, metabolic modules (defined as sub-pathways as annotated by KEGG24), and genes from human intestinal biopsies along the GI tract.

- But I am missing the results of negative/blank controls. Particularly, if you want to include the low abundant taxa into your output. What is the kitome of your workflow? [1, 2, 3, 4].

We appreciate the Reviewer bringing up this important concept and have attached an SI table containing the relative and absolute abundances of our blanks. We included three processing blanks, which were exposed to the clinic and treated with the entire MEM protocol. Each processing blank was sequenced analyzed with Quant-Seq to identify the microbial taxa introduced during processing and the absolute abundance of these taxon.

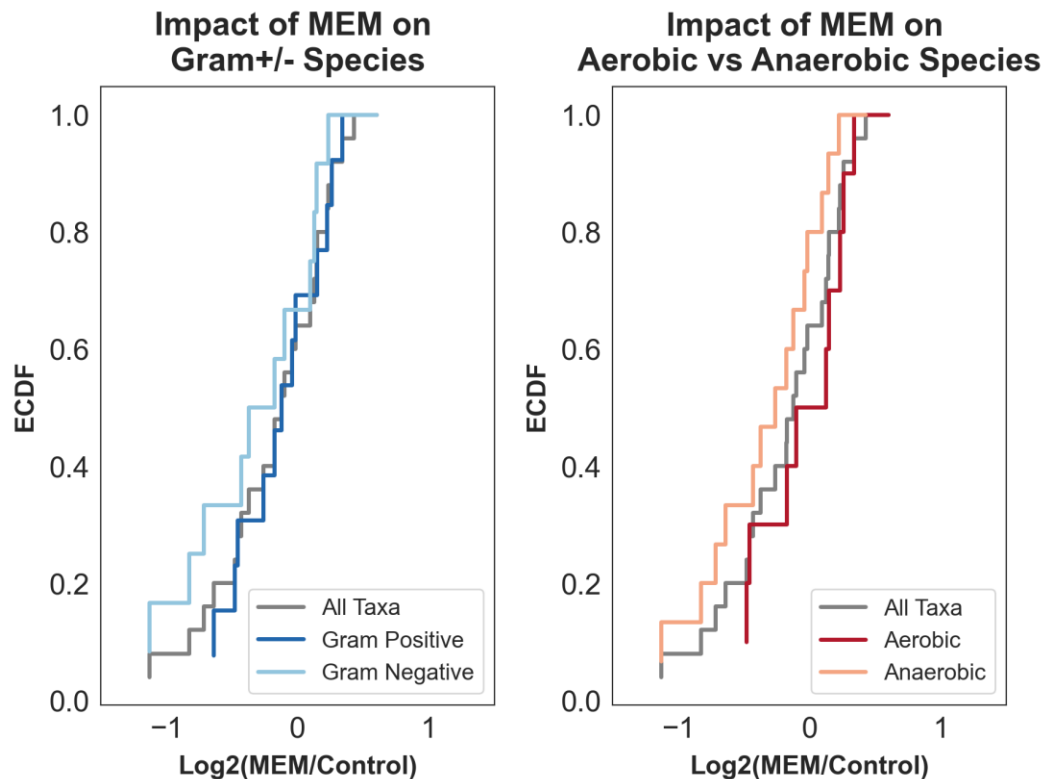
We also thank the Reviewer for sharing these excellent citations and have added them into the manuscript.

To determine how MEM performs on human intestinal biopsies, we recruited healthy participants undergoing routine colon cancer screenings via colonoscopy. From each of four participants, we obtained eight mucosal biopsies; four biopsies from each participant were assigned to the MEM-treatment group and four were untreated controls (Fig 3A). Due to concerns regarding contaminant DNA³⁴⁻³⁷ in samples with low bacterial loads, we also characterized the background bacterial signal associated with MEM and our processing methods through quantitative 16S rRNA gene sequencing of MEM processing blanks (see Methods and Tables S4-S5). MEM removed host DNA more than 2,000-fold across all 16 biopsies, with most biopsies having host levels comparable to a processing blank after MEM treatment. (Fig 3B).

34. Weyrich, L. S. et al. Laboratory contamination over time during low-biomass sample analysis. *Molecular ecology resources* 19, 982-996, doi:10.1111/1755-0998.13011 (2019).
35. Salter, S. J. et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC biology* 12, 87, doi:10.1186/s12915-014-0087-z (2014).
36. Velásquez-Mejía, E. P., de la Cuesta-Zuluaga, J. & Escobar, J. S. Impact of DNA extraction, sample dilution, and reagent contamination on 16S rRNA gene sequencing of human feces. *Applied microbiology and biotechnology* 102, 403-411, doi:10.1007/s00253-017-8583-z (2018).
37. Liu, Y., Elworth, R. A. L., Jochum, M. D., Aagaard, K. M. & Treangen, T. J. De novo identification of microbial contaminants in low microbial biomass microbiomes with Squeegie. *Nat Commun* 13, 6799, doi:10.1038/s41467-022-34409-z (2022).

• I am also missing positive controls with a known mixture and known concentration of Gram-positive and Gram-negatives, acid-fast bacteria, fungi etc. What is the effect of your proposed protocol on a broader range of community members?

We thank the Reviewer for this suggestion. We have added the below figure of the top 25 abundant species in human saliva showing the log₂ fold-change between MEM and control samples and classified each taxa as either Gram +/- or aerobic or anaerobic. There does not appear to be conserved bacterial losses following MEM treatment between each classification group although we note the highest losses are often from microbes classified as anaerobes.



Caption: ECDF of the log₂-fold change between MEM and untreated control human saliva samples for the top 25 most abundant species.

We have not characterized the effect of the MEM protocol on fungi. 16S rRNA gene sequencing does not permit characterization of fungi and we did not analyze shotgun sequencing data for fungal genes and/or taxa. Our manuscript states this limitation in the Discussion section:

We have only characterized the impact MEM has on bacteria and archaea. Future studies should understand whether MEM affects the mycobiome and virome.

• Although the results are positive (Figure 1D), the authors do not explain/discuss why MEM performs better than other methods. For instance, when using detergents to selectively lyse host cells, some bacterial cells may also be lysed due to differences in cell wall structures among bacterial species. In contrast, the big bead beating method selectively lyses cells much bigger than bacterial cells, reducing the level of bias between different bacterial species.

We have added the following statement to the Results reflecting why MEM performs better than other host-depletion methods:

We confirmed that MEM induced minimal losses in the microbial community; more than 90% of genera showed no significant difference in relative abundance between MEM and control samples (paired t-test, two-sided, $P=0.05$). Additionally, all taxa that were detected in the control samples were also detected in the MEM-treated samples, whereas MoLYsis and QIAamp host-depletion methods resulted in some taxa drop-out (Table S1). **Because MEM selectively lyses host cells based on cell size differences, this approach appears to introduce lower bacteria bias compared with chemical lysis alternatives (MoLYsis and QIAamp) where degree of lysis may differ based on bacterial cell wall/membrane structures (Fig S2).**

For each host lysis method, the authors should also determine if the differences between treatment/control are linked to taxonomy (e.g., replot Figure 3D as several sub-histogram, each for a different phylum).

We have reanalyzed 16S rRNA gene sequencing data at the genera level for individual bacterial phyla for mouse stool, human saliva, and human biopsies to characterize bacterial community changes in various phyla. The below plots have been added to the Supplement as Fig S2.

The phyla were chosen if at least ten different genera of that phylum were identified through sequencing. The dominant phyla in the human gut, Firmicutes and Bacteroidetes, appeared to be minimally affected by MEM treatment in all sample types. Actinobacter appeared to have higher losses following MEM treatment compared to other taxa in the human biopsies; however, the losses are not consistent across patients. For example, Collinsella, the Actinobacter genus that was most impacted by MEM, experienced roughly 4-fold to over 20-fold losses in relative abundance after MEM across different patients. Across these patients, Collinsella was found at the same relative abundances in the control biopsies. Therefore, it is difficult to determine if there are species/strain-level variations in how MEM affects microbial losses or if some of these microbes were already dead in one of the patients.

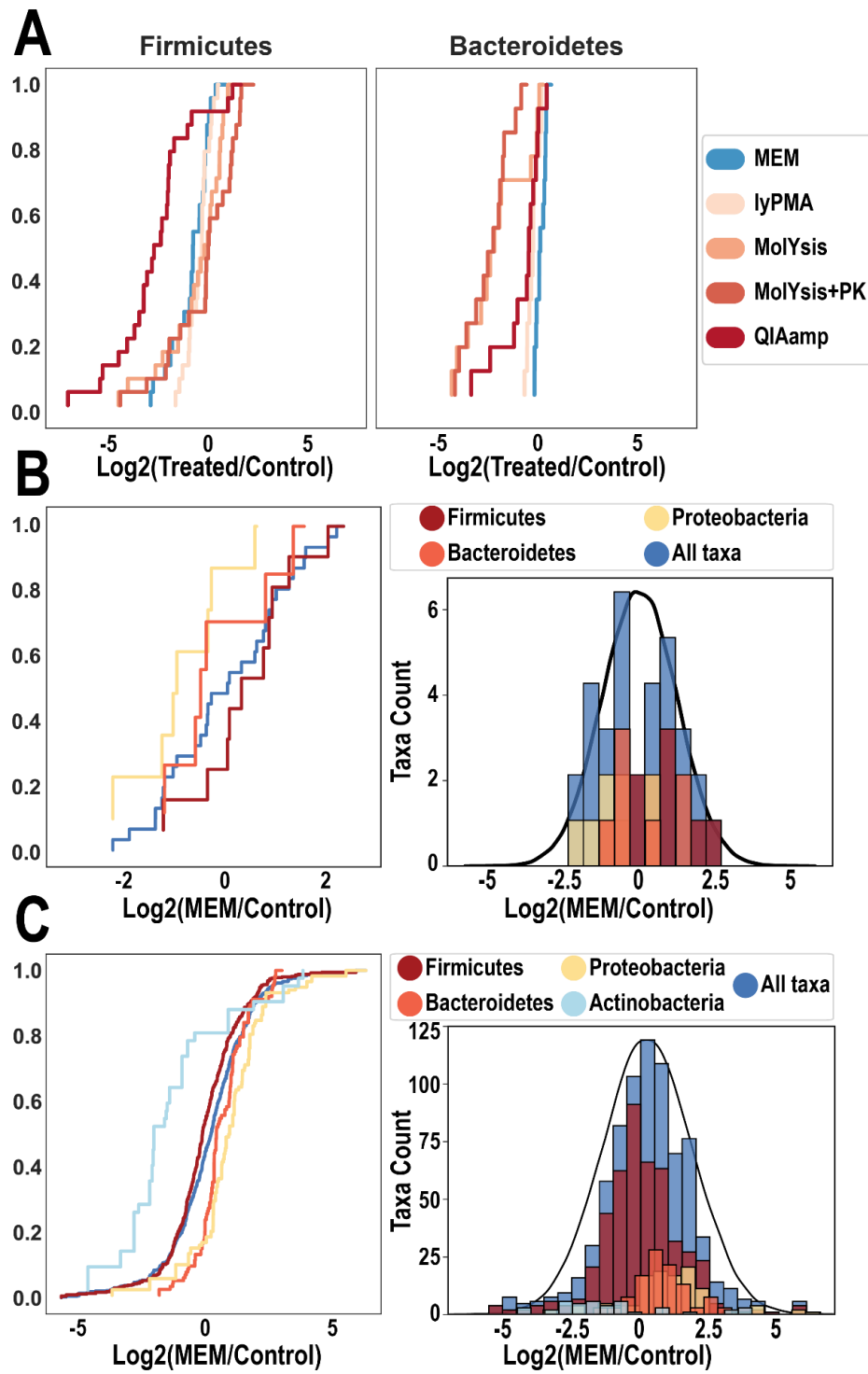


Fig S2: Log2 fold-change between relative abundance of genera within each phylum in treated and control samples from 16S rRNA gene sequencing data for A) mouse stool, B) human saliva, and C) human biopsies. The histograms in B-C) are overlaid with a normal distribution (black line).

- The estimation of bacterial DNA loss requires further clarification. The value displayed in Figure 1C does not directly reflect bacterial DNA loss, but rather represents the overall output of the host-DNA depletion methods and the subsequent bacterial DNA extraction process. The current conclusions seem to presume that the downstream bacterial DNA extraction method effectively retrieves all DNA molecules in a sample, which is not a valid assumption.

We agree with the Reviewer that we may not recover all DNA molecules in the sample; however, we have tried to standardize our recovery for all conditions by utilizing the same extraction kit and lysis protocol. We have clarified this in the Results:

Although we were unable to extract all DNA molecules in the samples, all samples were processed with the same extraction kit following host depletion to standardize extraction kit/lysis efficiency as much as possible. On homogenized stool samples, we observed similar losses in microbial recovery across all five host depletion protocols compared to a control, untreated sample (Fig 1C).

Additionally, certain pre-treatment steps during the host-DNA depletion process may even enhance the efficiency of downstream bacterial DNA extraction (Figure 2S).

We agree with the Reviewer that pre-treatment is likely increasing DNA recovery as shown clearest in Figure S2 in the MoLYsis condition. MoLYsis has an addition mutanolysin treatment step for increased bacterial lysis and shows increased bacterial yields. We have updated this point in the Results:

lyPMA appeared similarly effective at host removal as MEM but was difficult to use predictably because the stoichiometric nature of the method can result in large microbial losses when host levels are lower than expected (Fig S3).

Additionally, MolYsis showed increased bacterial recovery, likely due to the additional mutanolysis step.^{22,26}

22. Bruggeling, C. E. et al. Optimized bacterial DNA isolation method for microbiome analysis of human tissues. *Microbiologyopen* 10, e1191, doi:10.1002/mbo3.1191 (2021).

26. Cheng, W. Y. et al. High Sensitivity of Shotgun Metagenomic Sequencing in Colon Tissue Biopsy by Host DNA Depletion. *Genomics, proteomics & bioinformatics*, doi:10.1016/j.gpb.2022.09.003 (2022).

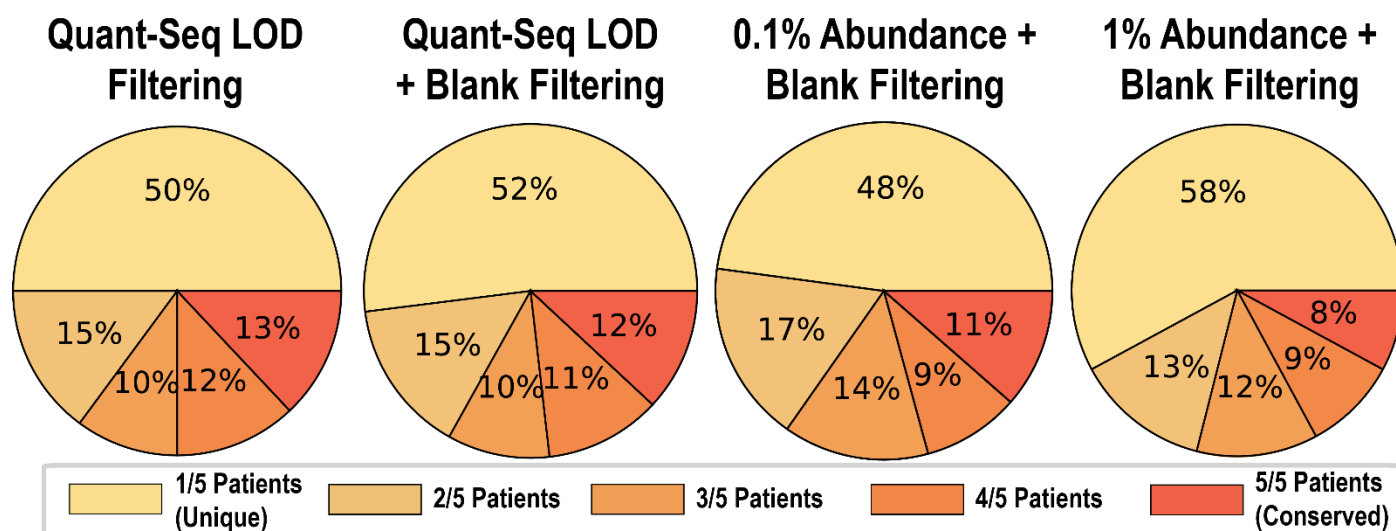
- “About half (91 of 187) of the microbial species identified were unique to an individual (Fig 4d).” This is a potentially significant finding, as species from stool are not this different between people, and the authors should elaborate further on this observation. What are the specific species that are unique to each individual? What are the species that are present in multiple individuals? Is this result merely due to the limited number of bacterial cells present in each biopsy sample?

We thank the Reviewer for posing these questions. We have provided a new supplementary table (see Table S7) showing the prevalence of all 187 microbial taxa so the reader can identify which species are classified as unique or conserved across individuals.

The result of roughly half microbial species being unique to an individual may be in part due to the limited number of bacterial reads present after shotgun sequencing. Due to the low sequencing depth used (9-45M reads), some samples were not sequenced to completion, therefore some taxa are identified as unique because they were not detected in any other participants. However, we found similar results from 16S rRNA gene sequencing at the genera level; 16S rRNA gene sequencing enabled detection of much lower abundance taxa because we were not read-depth limited in those low biomass biopsies.

It is possible that some uniqueness to each individual's microbiome could be explained by the effects of sequencing low biomass samples. However, to account for this, we considered a taxon to be a part of an individual's microbiome if the taxon was detected at least once. Given that we sequenced 12 biopsies from along the GI tract of each individual, even extremely low-abundance taxa are expected to be detected in at least one sample. We do recognize extremely low abundant taxa may still be missed across all 12 biopsies, but this is also the case for stool samples.

We also re-analyzed the 16S rRNA gene sequencing data with various relative abundance filters and found consistently the same results (see figure below). For the manuscript, we have redone this figure to show filtering at 0.1% relative abundance with blank filtering (Figure 4D), but as can be seen below, the selection of relative abundance filter does not change the conclusion that roughly half of the microbial taxa identified were unique to an individual.



It is difficult to determine the significance of this finding because our sample size is only five individuals. We do agree with the Reviewer that it would be interesting to study in the future on a larger scale and we anticipate that our microbial enrichment methodology will make this type of investigation feasible.

• The outcomes from the metabolic modules are greatly impacted by the current database and methodology, thus it's hard to assess the significance of the results. Additionally, for gut species, usually ~50% of genes can typically be annotated. Therefore, the fact that the overwhelming majority of genes from MEM metagenomes are unannotated raises concerns about the accuracy of the computational analysis (Figure 4A).

We thank the Reviewer for pointing out this potential inaccuracy. We agree that alignment of short reads to a reference catalog may overestimate the actual number of genes and pathways present in the sample. We have reanalyzed the short reads for pathway/gene analysis annotation using HUMAnN 3 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8096432/>). HUMAnN 3 has been extensively validated

on short reads and utilizes pangenomes of identified taxa from MetaPhlAn through marker gene analysis. We have updated all the pathway and gene analysis on short reads in Figure 4 based on the results from HUMAnN 3 (see below) and updated the Methods to reflect this.

We have made the following changes in the Methods section:

HUMAnN Pathway and Gene Alignment. Non-host read 1 and read 2 outputted from KneadData were concatenated and processed using the HUMAnN 3.0 workflow outlined in bioBakery (<https://github.com/biobakery/biobakery>) under default settings.⁶⁰ Taxonomic profiles obtained from MetaPhlan (see “Marker Gene Analyses”) were merged within patients and used as taxonomic inputs using the “--taxonomic-profile” flag in HUMAnN. Reported pathway abundances and gene abundances were normalized to relative abundances and concatenated.

The reanalysis with HUMAnN 3 does not change the overall conclusions for Fig 4A-B; MEM treated samples enable identification of many more pathways and genes compared to control samples. The main change is the number of genes detected across all samples is lower. This is expected as IGC represents a less curated but broader collection of genes compared to HUMAnN 3. Additionally, we did not require any minimum breadth of coverage when aligning to IGC, leading to an expected higher rate of false positives.

For Fig 4D, we note the metabolic modules from IGC appear more conserved than the pathways identified by HUMAnN. This is likely due to the annotation pipeline chosen as IGC utilizes KEGG metabolic modules and HUMAnN utilizes UniRef90 and MetaCyc annotations. In a paper comparing these two databases, authors note that “MetaCyc contains significantly more reactions and pathways than does KEGG” (Altman, et al. BMC Bioinformatics 14, 112 (2013)). This is further corroborated by the higher number of pathways detected after HUMAnN compared to IGC, even though we expect IGC to have a high percentage of false positives. Since IGC may be only annotating the more conserved, well documented pathways this would explain the lower percentage of conserved pathways when utilizing MetaCyc/HUMAnN.

For Fig 4G-H, we have decided to move the PCAs of the pathways and genes to the SI. These results are similar to our previous analysis with IGC.

We have updated the detected number of pathways/genes in the Results section. Additionally, we have changed the term “metabolic module” to “pathways” for consistency with the nomenclature used in HUMAnN.

We have made the following changes in the Results section:

We observed roughly a 100-fold increase in the number of organisms that could be detected, a 700-fold increase in the number of pathways detected, and over a 400-fold increase in the total number of genes detected in MEM-treated samples compared with the control samples (Fig 4A). When comparing only completed pathways, defined as above 90% complete, no complete pathways were detected in either of the control biopsies. An average of 1.5 (SD: 1.5) species and 728 (SD: 107) genes were detected in the control biopsies, whereas an average of 137.5 (SD: 21.5) species and 300,641 (SD: 6,922) genes were detected in the MEM-treated biopsies. MEM treatment enabled shotgun-sequencing classification of microbes down to a relative abundance of 0.005%, whereas in control biopsies a minimum relative abundance of 10% was required to detect microbes at similar sequencing depths. MEM-treated biopsies could detect genes down to a relative abundance of 10^{-10} , whereas in control biopsies genes could only be detected when present at a minimum relative abundance of 10^{-5} . We further found that MEM treatment improved reproducibility of detecting the most abundant genes. We compared the relative abundance of the top 5,000 most abundant genes between the two MEM-treated biopsies and the two control biopsies (Fig 4B). In the control biopsies, a high percentage of the genes (98%) were detected in only one sample, whereas for the MEM-treated samples only 3% of the genes were detected in one biological replicate but not the other.

Next, we tested whether MEM would enable characterization of microbial variation (at the taxon-, pathway-, and gene-level) cross sectionally across individuals and longitudinally across the GI tract of a single individual. Five participants undergoing annual colon cancer screening colonoscopy were sampled in four regions across the GI tract: terminal ileum, ascending colon, descending colon, and rectum. From each location, three biopsies were obtained resulting in a total of 12 biopsies per participant (Fig 4C). All biopsies were processed with MEM and the microbial profiles were characterized via 16S rRNA gene sequencing and shotgun sequencing at an average read depth of 25 million, producing an average of 2 million non-host reads (Fig 4D-F, Fig S4-S5). About half (91 of 187) of the microbial species identified were unique to an individual (Fig 4D). These unique species ranged in relative abundance from

10% to 0.01% (Fig S8). As was observed previously, **pathways and genes** appeared more conserved across participants compared with taxonomy (genera and species).^{38,39}

MEM enables detection of variation at the genus, species, metabolic modules, and gene levels in mucosal biopsies taken longitudinally across the GI tract

Whether mucosa-associated microbes vary along the GI tract has been challenging to determine due to the low number of microbial reads that could be recovered from mucosal biopsies. We first tested whether microbial variation between GI sites is present at the genus-level. For each participant sample, we used quantitative³³ 16S rRNA gene sequencing to quantify genus-level microbial changes longitudinally along the GI tract. Microbial taxa from the proximal colon (terminal ileum [TI] and ascending colon [AC]) and taxa from the distal colon (descending colon [DC] and rectum [R]) showed some clustering by location in most participants (Fig 4E). For each participant sample, we used shotgun sequencing to test whether the observed variation in taxa along the GI tract extended to the species, **pathway**, or gene levels. Clustering between the TI/AC vs the DC/R was seen in some participants across **species, namely in participants CT7, CT12, CT13, and CT14 (Fig 4F)**. There appeared to be minimal clustering between the TI/AC vs the DC/R at the pathway and gene-level (Fig S9). Additionally, there is high variation within regions for some individuals, which may be attributed to read depth limitations. For example, for one descending colon from CT13 no microbial marker genes were identified due to the minimal number of non-host reads so it is missing from the PCA plots (Fig 4F).

Shotgun sequencing of MEM-treated human intestinal biopsies enabled characterization of high- and low- abundance microbial species, **pathways**, and genes. This characterization documented longitudinal shifts in the mucosal microbiome along the lower human GI tract at the level of species, **pathways**, and genes.

We have updated Figure 4:

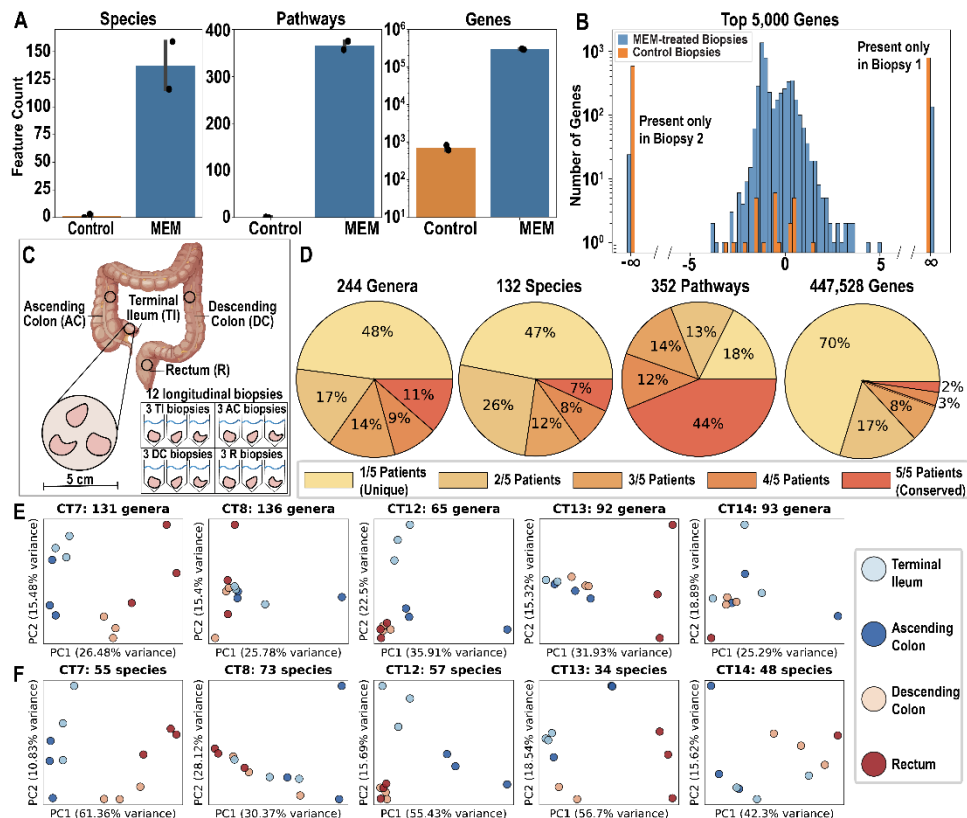


Figure 4: Shotgun sequencing of MEM-treated human intestinal biopsies. A) Four biopsies from participant CT18 (two MEM- treated and two control) were shotgun-sequenced and the number of microbial species, pathways, and genes identified in each sample were plotted (N=2; error bars are 95% CI). B) For the top 5,000 abundant genes, the log2 fold-change in relative abundances between the two MEM-treated biopsies and the two control biopsies were plotted. C) Sampling graphic for collection of 12 biopsies from each of 5 participants (3 biopsies taken from 4 separate regions of the GI tract). Biopsies within one region were sampled within one field of view (5-cm diameter). D) All 60 biopsies were processed with MEM, followed by 16S rRNA gene sequencing (for genera) and shotgun sequencing (for species, pathways, and genes). The number of features for genera, species, pathways, and genes were grouped based on whether they were present in at least one biopsy sample from only one participant (1/5), two participants (2/5), three participants (3/5), four participants (4/5), or from all participants (5/5). Analysis was performed on annotated genes, and separately on all genes (annotated + unannotated). E-F) PCA was performed on all 60 longitudinal samples grouped by participant (CT7, CT8, CT12, CT13, CT14). E) PCA on relative abundance of

16S rRNA gene sequencing genera assignments. F) PCA on relative abundance of shotgun-sequencing species assignments.

We have added Figure S9 to the Supplement:

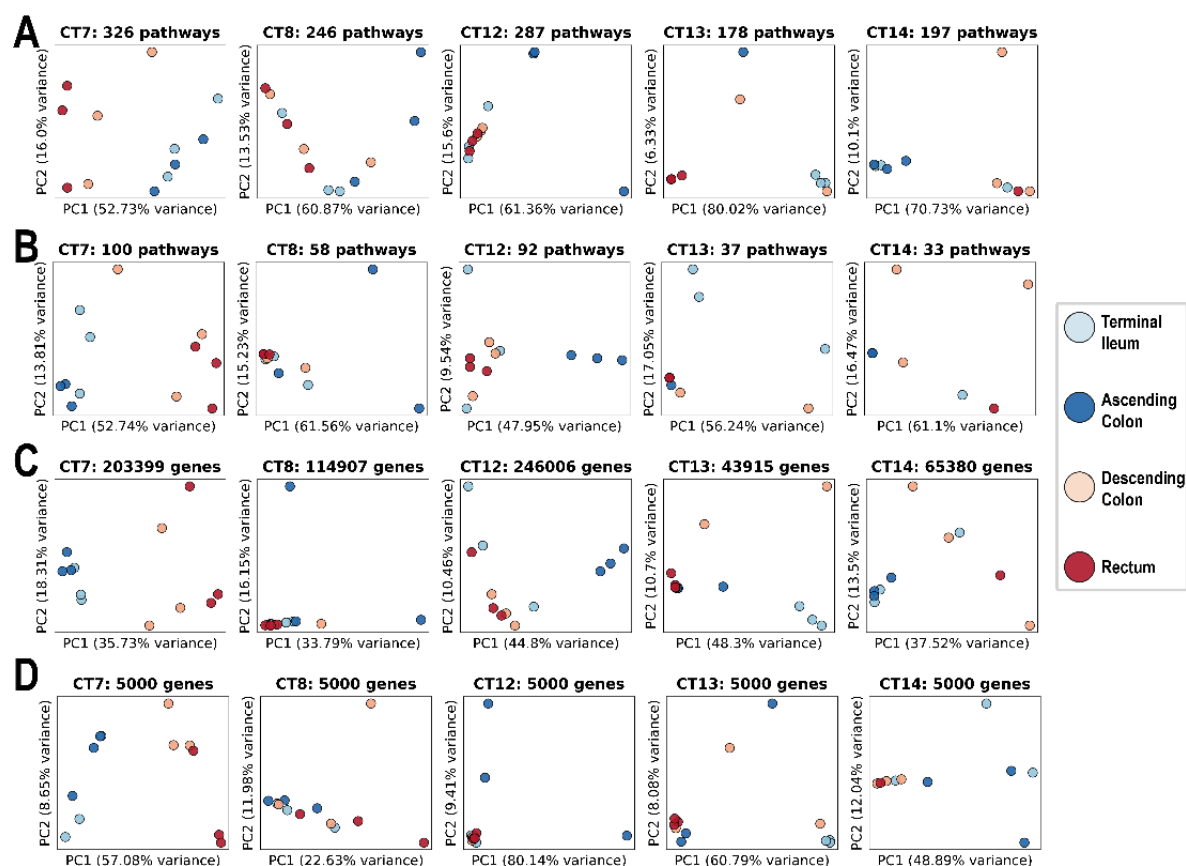


Fig S9: Longitudinal variation at the pathways and gene-level. PCA analysis was performed on all 60 longitudinal samples grouped by participant (CT7, CT8, CT12, CT13, and CT14). Shotgun-sequencing data was annotated for pathways and genes through HUMAnN 3. A) PCA on relative abundance of all pathways. B) PCA on relative abundance of completed pathways (defined as above 90% of modules being present). C) PCA on relative abundance of all genes. D) PCA on relative abundance of the top 5,000 most abundant genes in each participant.

- The authors assess the quality of MAGs using completeness and contamination, which is insufficient in many ways. How does the MAG size compare to the typical genome size of isolates from that species? Are the MAGs containing 16S, 23S, and 5S sequences? (Also check the MIMAG reference: <https://doi.org/10.1038/nbt.3893>.)

We are in complete agreement with the Reviewer that the completion and redundancy estimates based on SCGs alone are not suitable metrics to discuss the actual quality of MAGs. We have even reported previously that MAGs that are fully compatible with MIMAG quality criteria may suffer from significant contamination (<https://genome.cshlp.org/content/30/3/315.full>). We also thank the Reviewer for pointing out the missing information in our reporting for others to be able to judge the quality of our MAGs; indeed, the number of contigs and the size of a given MAG compared with its evolutionary homologs is one of the best ways to gain confidence in its quality. We now have expanded the description of our MAGs in the Supplementary Table 9, and for each MAG we report (1) the number of contigs, (2) N50, (3) number of tRNA and rRNA genes, (4) MAG length, (5) the average length of genomes in the Genome Taxonomy Database (GTDB) that match the taxonomic classification of a given MAG, and (6) the ratio between the size of each MAG and the average size of the genomes with the same taxonomy. We are happy to report that we found a very high correspondence between the lengths of our MAGs and their best matching reference genomes in the GTDB ($R^2 = 0.8057728$, $p = 2.2e-16$). We updated the main text to mention the correspondence between MAGs to what is known about these taxa.

We have updated the following text in the Results:

Next, to assess whether any of these MAGs were contaminated, we performed taxonomic classification on each genome.^{40,41} With a threshold of 95% average nucleotide identity (ANI), 33 MAGs were successfully classified. We compared the size of each classified MAG with the matching reference genomes in GTDB and found high agreement with current microbial databases ($R^2 = 0.8057728$, $P = 2.2 \times 10^{-16}$) indicating that the MAGs constructed from MEM-treated samples were not artifacts (Table S9). One *Fusobacterium* MAG matched closely with a published fecal-derived MAG at 86.85% ANI but GTDB was unable to assign species-level taxonomy (Fig S10).

- Detection of SNVs (SNPs) by comparing MAGs is not a reliable method as the assembly process often introduces numerous errors. The conventional way to detect SNVs is to use sequence alignment tools such as bowtie2 and samtools. The authors should refer to the many papers calling SNVs from metagenomic datasets (e.g., doi.org/10.1371/journal.pbio.3000102). The use of the fixation index requires certain assumptions to be met, and it is unclear if those assumptions are satisfied in this case. The authors should either clarify the concept of a fixation index and why it is appropriate to use here, or opt for a simpler metric, such as counting the number of SNVs between samples, which could still lead to the same conclusion.

We agree with the reviewer that in most cases the population genetics approaches that utilize single-nucleotide polymorphisms (SNPs) found on MAGs will suffer from increased number of variants due to biases associated with assembly. However, we would like to note that in our study we recover single-nucleotide variants (SNVs) by using MAGs as reference sequences to characterize the structure of matching naturally occurring populations in distinct samples through metagenomic read recruitment, rather than comparing MAGs to one another. This is a well-established approach to study microbial population genetics (<https://www.science.org/doi/10.1126/sciadv.abq4632>), and the method we use to estimate similarities between distinct metagenomes, fixation index, is designed particularly to operate on such data (<https://doi.org/10.1038/nature11711>) as it allows for variable positions to have multiple competing alleles, which is often the case in naturally occurring populations. We now have updated the text to clarify this point and avoid misunderstandings.

We have updated the following text in the Results:

SNVs detectable across locations in the lower GI tract within an individual

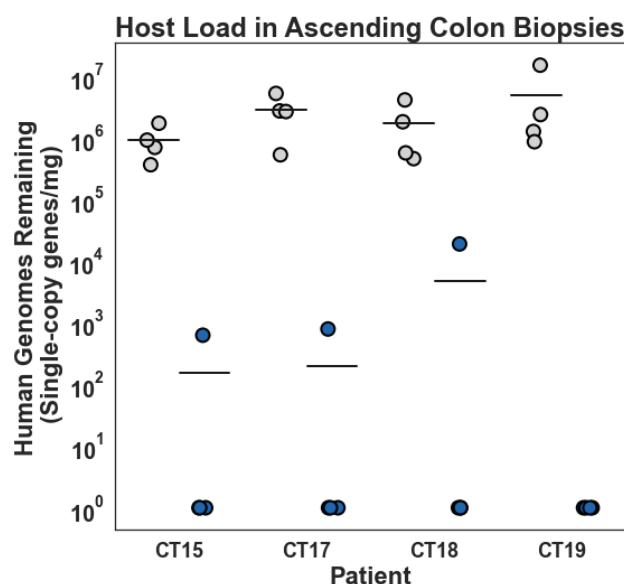
Finally, we investigated whether MEM treatment could enable studies of microbial population genetics in low- biomass samples through single-nucleotide variants (SNVs) as a result of the increased depth of coverage. For this analysis, we analyzed MAGs from patient CT12 as reference genomes and mapped the paired-end reads from the terminal ileum and descending colon from CT12 onto these assembled genomes. Six MAGs had a mean coverage above 50X across all 6 samples (3 terminal ileum and 3 descending colon) and were selected for subsequent SNV analysis (Fig S14). SNV profiles were generated from the paired-end reads of each sample by comparing them with the reference sequence (MAG). We investigated whether PCR errors may be responsible for some of the SNVs observed in our data by preparing libraries for an additional three technical replicates from a single terminal ileum biopsy (Fig 6b), with the expectation that differences in the SNV profiles of the technical replicates should be minimal.

- The method used for normalizing host depletion is not described (e.g., see Figures 1E, 1F, 1G).

According to the method section, each sample may have different starting weights or volumes, resulting in varying amounts of host DNA. Normalization criteria may also differ based on the type of sample. It is unclear whether the number of genomes remaining is based on a specific volume of DNA extract, the entire sample, or some other definition.

We have updated the captions for Figure 1E-G and 3B (see below). The method section discussing the various starting weights was referring to human biopsies (data shown in Figure 3B). We did not normalize this data as the biopsies varied in weight by about an order of magnitude (0.2-4.8 mg) while the depletion is over 3 orders of magnitude (1,000-fold). We also ensured even splitting of the various biopsy sizes into the two different treatment groups (see Methods: “After samples were transferred to the lab, biopsies from the same passage were split into control and depleted groups for a total of 4 biopsies per condition with evenly sized biopsies present in each group”). Therefore, any noise introduced by variation in the biopsy weight is negligible compared to the difference between depleted and non-depleted. We have included the normalized host load data to demonstrate that the conclusions are largely unaffected (see below), but we personally find the non-normalized to be clearer as we can draw a continuous LOB line across samples. However, if the Reviewer feels strongly about showing the normalized data, we are happy to make that change.

Here is an example of Figure 3B normalized to biopsy mass. Samples that were below LOB were set to a load of 0 copies/mg.



For the validation done in Figure 1E and 1F, the samples are homogenized prior to splitting (saliva is vortexed between samples to keep well mixed and mucosal scrapings are pipetted up and down in PBS and vortexed between samples). For the hard tissue sample used in 1G, we are unable to homogenize this sample type prior to splitting. However, there is variability between sampling replicates compared to the depletion results (I.e. sampling replicates may differ by a maximum of 2-fold while depletion results in a 10^3 fold change). We do not expect the small differences between samplings replicates to affect the overall conclusion.

We have made the following changes to the Figure 1 and Figure 3 captions:

Figure 1: Comparison of the performance of the microbial-enrichment methodology (MEM) with published host-depletion methods. ... E-G) Remaining host DNA was quantified through ddPCR of a single-copy host specific primer (see Methods). Reported genomes remaining refers to the abundance of this single-copy gene present in 1 μ L of elution. E) Remaining human genomes in fresh human saliva were quantified after treatment with each host-depletion method and in untreated controls (N=3). F) Host-depletion methods were tested on mouse intestinal mucosal scrapings as a representative of soft tissue and remaining mouse genomes were quantified (N=3; biologic

replicates from one mouse). G) Host-depletion methods were tested on rat colonic sections as a representative hard tissue (including connective tissue, muscle, and mucosa) and remaining rat genomes were quantified (N=3; biologic replicates from one rat).

Figure 3: Analysis of microbial enrichment in paired human intestinal biopsies processed with and without MEM A) Sampling graphic illustrates collection from four participants (CT15, CT17, CT18, and CT19) each with eight ascending colon biopsies. Four biopsies from each participant were MEM-treated and four were untreated controls. B) Host DNA was quantified for each biopsy using ddPCR of a single-copy host primer. Human genomes remaining refers to the abundance of this single-copy gene present in 1 μ L of elution (* indicate measurement was below limit of blank [LoB] defined as $LoB = mean_{blank} + 1.645[SD_{blank}]$ based on three processing blanks).

- The use of detection score might raise the concern that the MAGs may include contigs that are contaminations and detection score cannot tell if that's the case. The authors should use a metric like absolute of normalized sequencing depth or provide more justification for using detection score.

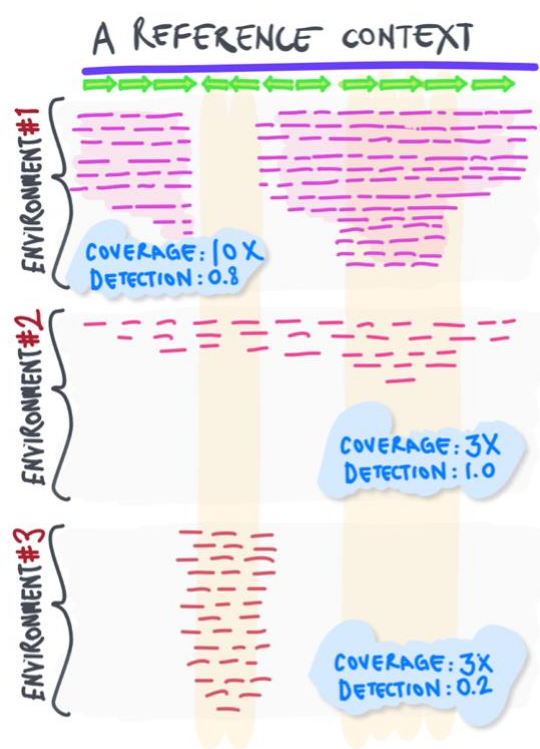
We thank the Reviewer for bringing up this suggestion. First, we apologize for any confusion the term “detection score” may have caused as the more correct terminology is “detection.” We have updated the manuscript to use the term “detection” to promote continuity with previously published work.

We would also like to note that the detection metric is not used when binning and generating MAGs. MAGs are binned based on differential coverage across samples and sequence composition, as has been described previously (Kang DD, et al. PeerJ. 2019;7:e7359. 2019, Alneberg, J. et al. Nat Methods 11, 1144- 1146, 2014,

Yang Young Lu et al., *Bioinformatics*, Volume 33, Issue 6, 2017). Detection was only used when determining whether a gene or genome was present within the sample. We agree detection does not provide evidence whether some contigs are contaminating.

Detection is also referred to as the breadth of coverage (i.e. the proportion of nucleotides in a given reference sequence that are covered by at least one short read in a given metagenome). Detection is an extremely effective way to be able to discuss the presence of a given population in a given environment, independent of read coverage, and by avoiding false positives due to non-specific read recruitment. Non-

specific read recruitment is a common phenomenon in metagenomic read recruitment since the metagenomes are almost always more complex and diverse than the reference sequence sets used for read recruitment, and as a result some reference sequences recruit reads from non-target populations due to sequence homologies. The following drawing illustrates the utility of detection in such scenarios:



Thus, the detection metric is essential to be able to discuss the presence of a given population represented by a MAG in our datasets. Detection essentially reflects the percentage of a MAG that has at least 1X coverage in a given environment.

We have now updated the text regarding its utility to avoid any misunderstandings and thank the Reviewer for bringing this up and have made the following changes in the Results:

For the 34 high-quality MAGs, we computed detection, which reports the proportion of nucleotides in a given reference sequence that are covered by at least one short read in a given metagenome. Thus, detection is an extremely effective way to be able to discuss the presence of a given population in a given sample, independent of read coverage, and by avoiding false positives due to non-specific read recruitment. To confirm that the MAGs reconstructed from

MEM-treated samples were accurate representations of the untreated biopsies, we assessed the uniformity of coverage of the control reads when mapped back onto the MAGs (Fig. 5C). To perform this analysis, we chose a MAG resolved to *Alistipes putredinis*, a known gut microbe that had the highest detection in the control samples. Control samples showed an even distribution of reads among the 29 contigs present, indicating that this MAG was also present in the control samples, but sequencing depth limitations prevented the reconstruction of a genome. Overall, we observed a higher detection of all 34 high-quality MAGs in MEM-treated samples compared to control samples (Fig 5D).

For saliva, we observed a high correlation in the relative abundances of microbial taxa between the control sample and both the MEM and MEM+DTT treatments (Fig 2d). The correlation was less pronounced for low-abundance taxa with enrichment of specific species in MEM-treated samples

• I would like to see the negative controls here. What is the background noise of MEM treatment? Is the observed enrichment of specific low abundance species signal or noise?

We appreciate this suggestion by the Reviewer. We have included an SI table (Table S4-5) containing the relative and absolute abundances of our 16S rRNA gene sequenced processing blanks. We will also add the below figure to the SI (Fig S5) that demonstrates bacterial loads of the processing blanks compared with the bacterial loads in the saliva samples. This figure shows that the bacterial load in the blanks is roughly 6 orders of magnitude lower than in the saliva samples. Thus, any contaminating taxa are below 10^{-4} relative abundance and would be filtered out during analysis of the saliva samples.

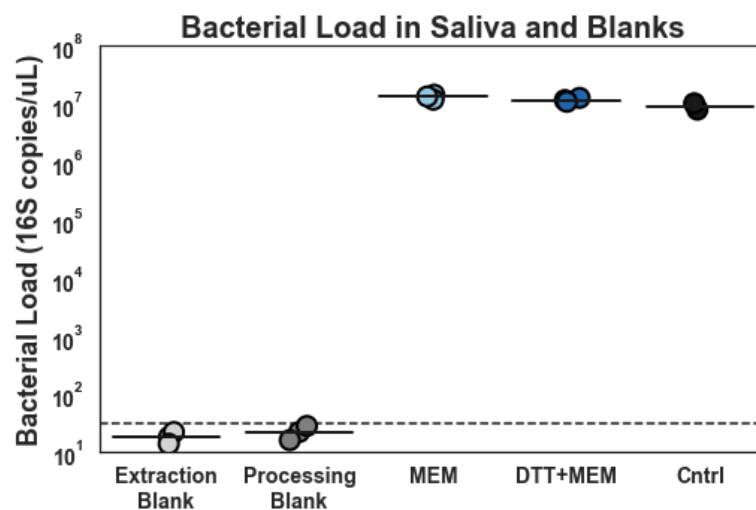


Figure S5: Bacterial load present in saliva and extraction/processing blanks. Bacterial load measurements for saliva samples that were shotgun sequenced compared to processing and extraction blanks. Dashed black line represents the measurement LOB.

To further confirm that the observed enrichment of specific species in the MEM samples was not due to contamination, we reanalyzed the shotgun sequencing data for saliva and identified 17 genera that were consistently found in MEM-treated samples (DTT+MEM and MEM-only samples) but were absent from control samples. Then we compared the 16S rRNA gene sequencing data of these same samples. Of the 17 genera, 6 were present in control and MEM-treated samples at low abundances (0.05–0.5% relative abundance). The remaining genera were not found in control or MEM treated samples, potentially due to nomenclature mismatch between the SILVA and MetaPhlAn databases. We also checked whether any of these genera were found in 16S rRNA gene sequencing data of MEM processing blanks. All blanks were negative for all 17 genera.

We chose to use 16S rRNA gene sequencing for this investigation due to the difficulties associated with obtaining enough biomass to shotgun sequencing either extraction blanks or the clinical processing blanks.

We also investigated the identities of these genera and found that all 17 genera are known human commensals and the majority are associated with the oral cavity. Therefore, we conclude these 17 enriched taxa are likely not contaminants. An SI table containing all the species and relative abundances identified from shotgun sequencing for both MEM, DTT+MEM, and Control saliva samples will be uploaded (Table S3).

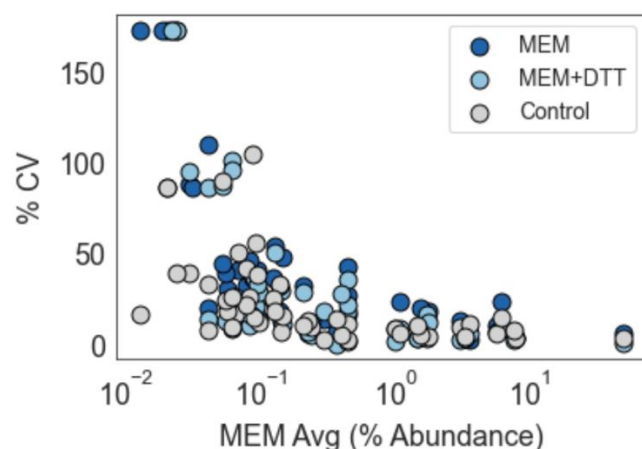
MEM-treated saliva samples had lower CV (50% CV, 95% CI), indicating better replicability compared with untreated controls.

- Again, if there is a persisting contamination background signal of certain taxa in MEM-treated samples, the replicability increases because the background signal has a high recognition value.

We appreciate this point made by the Reviewer. As we discussed in the previous comment, the bacterial load detected in saliva was about 6 orders of magnitude higher than the load detected in any of the blanks.

Since the taxa showing lower CV in MEM-treated samples were 0.01-0.1% relative abundance, these taxa are likely not originating from the kitome.

We do agree with the Reviewer that we had not fully addressed why MEM-treated samples had a lower CV. This may be due to increased bacterial lysis and homogenization of samples due to additional pre-processing from the MEM-treatment or it may be due to the increase in non-host reads. To clarify this, we analyzed paired 16S rRNA gene sequencing data of these samples, as 16S rRNA gene sequencing will not have the same read-depth limitations for the control saliva samples. Replotting the %CV of 16S rRNA gene sequencing of saliva samples reveals a similar %CV for all treatment conditions for low abundance taxa. This implies the high reliability seen in the control samples from shotgun sequencing is likely a repercussion of the fewer number of non-host reads.



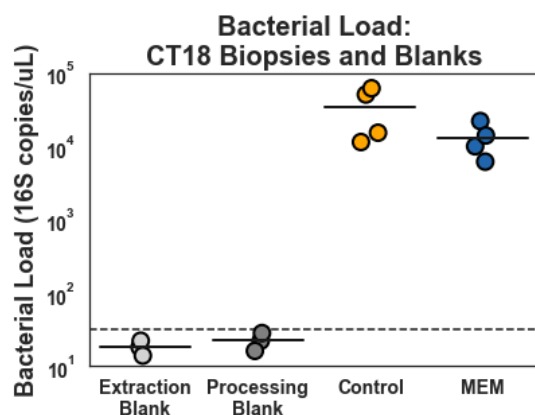
Caption: Coefficient of variation was plotted against relative genera abundance from 16S rRNA gene sequencing and colored based on treatment types. Each point represents a genera.

we shotgun-sequenced paired control and MEM-treated biopsies from CT18 at a depth of above 100 million reads (N=2 for each condition) (Fig 3a). We observed roughly a 100-fold increase in the number of organisms that could be detected

- Again, I would like to see the blank, negative controls here for MEM-treated biopsies and controls. Detecting more taxa is not necessarily beneficial, if it's just more contamination.

We have uploaded the 16S rRNA gene sequencing data for processing blanks (Table S4-5). Bacterial loads in the processing blanks were ~3 orders of magnitude lower than the loads in participant CT18's

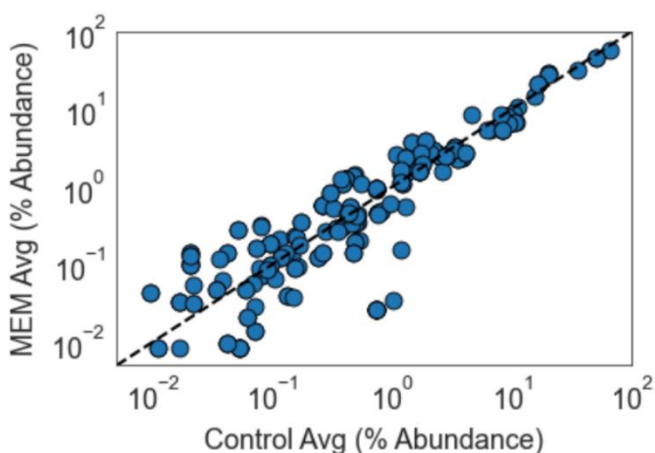
MEM-treated samples, indicating that any contaminating taxa introduced during MEM processing must be present at low abundances (see figure below).



Caption: Bacterial load in processing blanks and biopsies for CT18. Bacterial load was measured from the 100uL of extraction elution using 16S rRNA gene primers through digital PCR.

However, we noted that the bacterial loads in some intestinal biopsies were quite low, and comparable to the processing blanks. As a result, we also performed blank filtering on 16S rRNA gene sequencing data from intestinal biopsies using the 3 processing blanks (please see the updated Methods below). If a taxon was detected in any of the processing blanks at a higher absolute abundance compared to any of the biopsy samples, the entire taxon was removed from downstream processing. After blank filtering, over 93% of genera remained. We agree with the Reviewer that some contaminating signal may be introduced from MEM treatment; however, the majority of the bacterial signal detected is not due to background contamination. Please see below the updated language in the Methods and Results describing the blank filtering.

To confirm that taxa were not being added to the MEM-treated sample, we analyzed 16S rRNA gene sequencing of MEM-treated and control biopsies from participant CT18. If significantly more taxa were detected in MEM-treated biopsies this may indicate that significant background contamination was introduced by MEM-treatment. However, we identified nearly the same number of taxa in MEM-treated and control biopsies with 245 taxa detected across all four MEM-treated biopsies and 243 taxa detected across all four control biopsies from CT18. Additionally, when plotting the relative abundance in MEM-treated vs control biopsies, we see high concordance in taxon abundances between the two conditions ($R^2 = 0.958$).



Caption: The average relative abundance obtained from 16S rRNA gene sequencing of the 4 control biopsies and the 4 MEM-treated biopsies from CT18 were plotted against each other for each taxon. Overlaid in black is a 1:1 correlation.

We performed contamination filtering on 16S rRNA gene sequencing data due to the difficulty in generating enough biomass to shotgun sequence blanks. We agree with the Reviewer that this data would be ideal but believe the 16S rRNA gene sequencing data of the 3 processing blanks provides sufficient characterization of the minimal contaminating signal MEM may introduce. Additionally, all samples were processed in the same manner for shotgun sequencing (I.e. both control and MEM-treated biopsies were exposed to the same reagents and processing environment). Therefore, if any background signal was introduced during the shotgun sequencing library prep, it should be consistent in both MEM and control biopsies.

We have made the following changes in the Methods section:

Longitudinal Sampling of the Human Intestinal Tract (Figure 4C)

For longitudinal sampling, a total of five participants were sampled 12 times from 4 different locations during a routine colonoscopy. The 4 locations sampled were the terminal ileum, ascending colon (designated as 10 cm distal to the ileocecal valve), descending colon, and rectum. From a single field of view (5 cm diameter) from each location, 3 biopsies were collected in one passage with 2.8 mm biopsy forceps and stored dry on ice in a microfuge. **For participant CT14, only one rectal sample was obtained.** On average, biopsies were 2.5 mg with a minimum size of 0.1 mg and a maximum of 5.9 mg. All biopsies were then processed

individually in the laboratory at University of Chicago (see “MEM”) before shipment on dry ice to Caltech for DNA extraction. Time between specimen collection and processing ranged from 10 min to 52 min. Samples were processed individually in the laboratory at University of Chicago (see “MEM”) before shipment on dry ice to Caltech for DNA extraction. Additionally, three microfuge tubes of 400 µL saline were opened and closed in the laboratory to serve as clinical processing blanks.

Quant-Seq. Microbial characterization and quantification was performed using the quantitative sequencing (“Quant-Seq”) pipeline we have described previously.³³ Due to the low bacterial loads present in intestinal biopsies, Quant-seq was also performed on three MEM processing blanks. If a taxon was detected at a higher absolute abundance in any of the processing blanks compared to the intestinal biopsies, the taxon was removed from downstream analysis.

We have made the following changes to the Results:

To determine how MEM performs on human intestinal biopsies, we recruited healthy participants undergoing routine colon cancer screenings via colonoscopy. From each of four participants, we obtained eight mucosal biopsies; four biopsies from each participant were assigned to the MEM-treatment group and four were untreated controls (Fig 3a). Due to concerns regarding contaminant DNA³⁴⁻³⁷ in samples with low bacterial loads, we also characterized the background bacterial signal associated with MEM and our processing methods through quantitative 16S rRNA gene sequencing of MEM processing blanks (see Methods and Tables S5-S6). MEM removed host DNA more than 2,000-fold across all 16 biopsies, with most biopsies having host levels comparable to a processing blank after MEM treatment. (Fig 3B).

To determine how MEM affects the human intestinal microbiome at a community level, we first performed quantitative 16S rRNA gene sequencing.³³ Roughly 93% of genera remained in MEM-treated and control biopsies after computationally removing taxa found at higher absolute abundances in the blanks, giving us confidence most detected taxa were not background contaminants. To further confirm that MEM did not introduce additional contamination, we found strong agreement between taxon abundances in MEM and control biopsies (Table S6).

PCA of sequencing results showed that any differences in microbial relative abundances introduced by MEM were less than the differences observed between participants (Fig 3C). Analysis of sequencing results revealed minimal changes in relative abundances of most taxa after MEM treatment, with roughly 88% of taxa having no significant differences in relative abundances from the controls (Mann Whitney U test, two-sided $P=0.05$). For taxa present at greater than 1% relative abundance, more than 95% of taxa had no significant differences between MEM and control samples. The log₂-fold difference in taxa between control and MEM-treated samples approximated a normal distribution (Kolmogorov–Smirnov test against normal distribution, statistic=0.074, $P=0.11$) (Fig 3D). Furthermore, there was a linear correlation in relative taxon abundances between the control and MEM-treated samples (Fig 3E). The speed and simplicity of MEM enables over 1,000-fold host removal while introducing minimal biases in microbial relative abundances when used in a clinical setting on human intestinal biopsies.

Figure 4: Shotgun sequencing of MEM-treated human intestinal biopsies

- Please show and quantify whether individual patterns persist as a function of time instead of showing all the individual PCA plots. Why is there so much variance in some individuals (Figure 4), e.g., for CT13, figure 4G, one descending colon sample is found on the complete opposite site of the x-axis compared to the other two descending colon samples with a variance of about 60%. Also, I only detect two descending colon samples for CT13 (figure 4F). Where is the third one? For CT14, 4F I only detect 1 rectum sample (color red). Where are the others? Why is there sometimes more variance (PC1) on highly compressed module than gene level (CT8, CT12, CT13)?

We thank the Reviewer for bringing up these great points. For two of the PCA plots mentioned (previously figure 4G-H) plotting pathways and genes, we re-analyzed this data based the Reviewer's comments regarding our metabolic module and gene annotation method. The revised PCA plots of pathways and genes using HUMAnN 3 to annotate will be added to the SI (Fig S9) and can be seen below. We do note there are still high variation in some individuals, namely CT13. This may be largely attributed to the low microbial read counts obtained from these samples, specifically in the descending colon. As the Reviewer pointed out in figure 4F there were only two descending colon samples. This is due to the low microbial read counts for this sample, which resulted in MetaPhlAn identifying no microbial marker genes. As can be seen from the SI (Fig S5), the bacterial loads for CT13 were quite low, specifically in the descending colon and rectum.

We also thank the Reviewer for noting the single rectal sample for CT14. During collection, we were unable to obtain three biopsies from this region so there is only one rectal sample. We have updated the Methods section to clarify this.

We are unsure what data the Reviewer is requesting “as a function of time”. We would like to clarify that each participant was only sampled at a single time point. Longitudinal sampling refers to longitudinal spatially rather than temporally. We agree this would be interesting data to obtain in the future.

We have made the following changes in the Methods section:

Longitudinal Sampling of the Human Intestinal Tract (Figure 4C)

For longitudinal sampling, a total of five participants were sampled 12 times from 4 different locations during a routine colonoscopy. The 4 locations sampled were the terminal ileum, ascending colon (designated as 10 cm distal to the ileocecal valve), descending colon, and rectum. From a single field of view (5 cm diameter) from each location, 3 biopsies were collected in one passage with 2.8 mm biopsy forceps and stored dry on ice in a microfuge. **For participant CT14, only one rectal sample was obtained.** On average, biopsies were 2.5 mg with a minimum size of 0.1 mg and a maximum of 5.9 mg. All biopsies were then processed individually in the laboratory at University of Chicago (see “MEM”) before shipment on dry ice to Caltech for DNA extraction. Time between specimen collection and processing ranged from 10 min to 52 min.

We have made the following changes in the Results section:

For each participant sample, we used shotgun sequencing to test whether the observed variation in taxa along the GI tract extended to the species, **pathway**, or gene levels. Clustering between the TI/AC vs the DC/R was seen in some participants across **species, namely in participants CT7, CT12, CT13, and CT14.** **There appeared to be minimal clustering between the TI/AC vs the DC/R at the pathway and gene-level (Fig S9).** Additionally, there is high variation within regions for some individuals, that may be attributed to read depth limitations. For example, for one DC sample from CT13 no microbial marker genes were identified due to the minimal number of non-host reads so it is missing from the PCA plots (Fig 4F).

We have added Figure S9 into the Supplement:

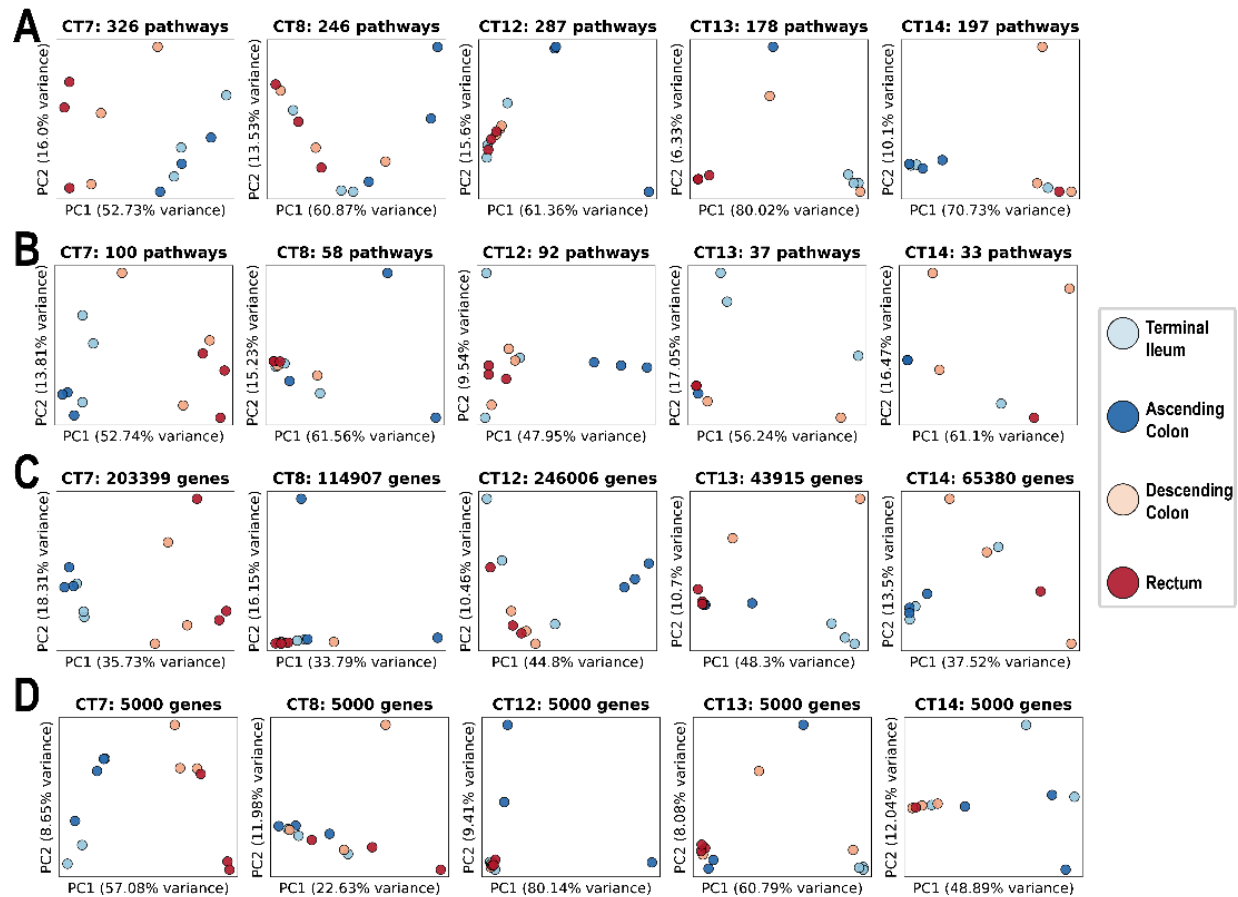


Fig S9: Longitudinal variation at the pathways and gene-level. PCA analysis was performed on all 60 longitudinal samples grouped by participant (CT7, CT8, CT12, CT13, and CT14). Shotgun-sequencing data was annotated for pathways and genes through HUMAnN 3. A) PCA on relative abundance of all pathways. B) PCA on relative abundance of completed pathways (defined as above 90% of modules being present). C) PCA on relative abundance of all genes. D) PCA on relative abundance of the top 5,000 most abundant genes in each participant.

We have updated Figure 4:

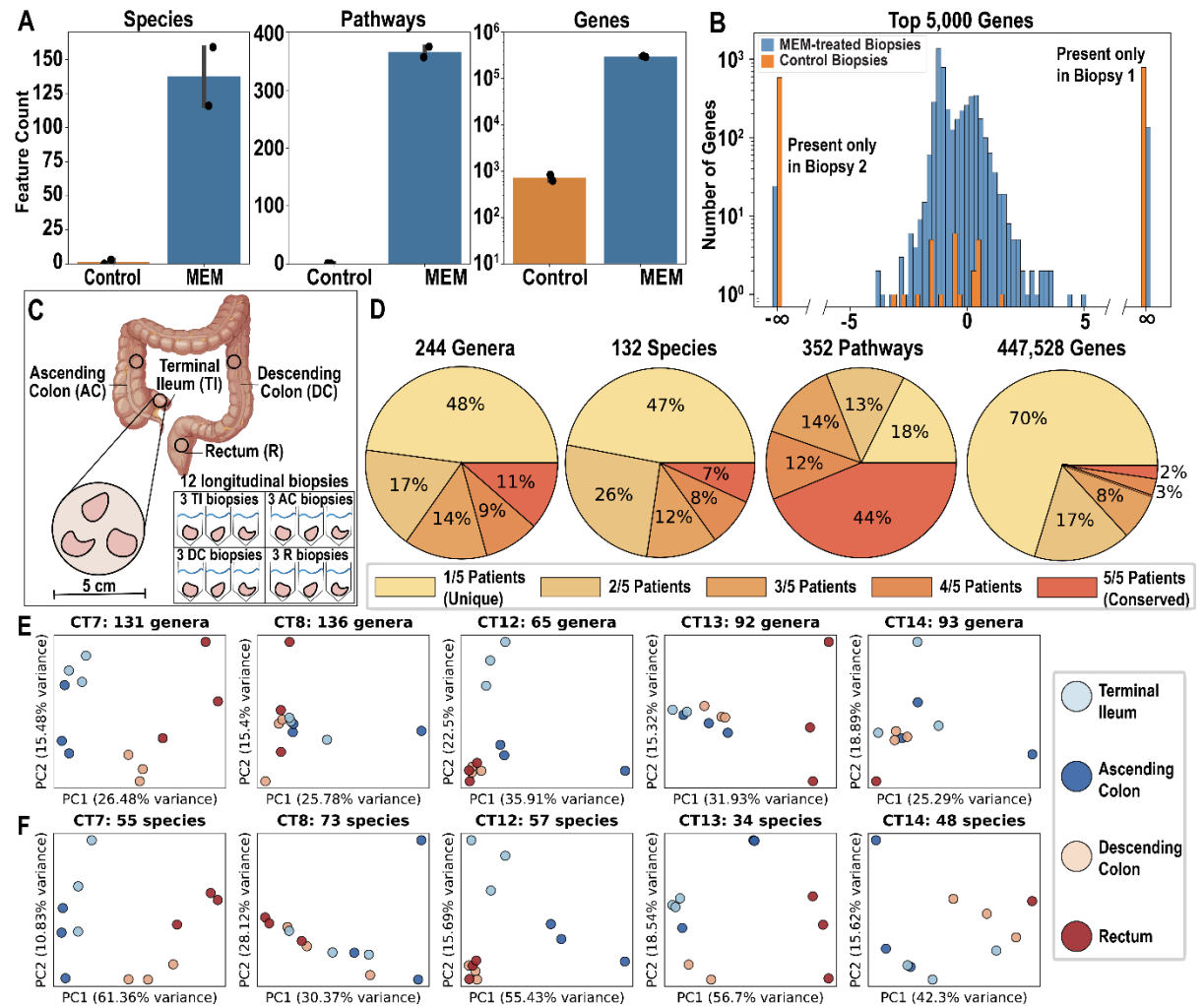


Figure 4: Shotgun sequencing of MEM-treated human intestinal biopsies. A) Four biopsies from participant CT18 (two MEM- treated and two control) were shotgun-sequenced and the number of microbial species, pathways, and genes identified in each sample were plotted (N=2; error bars are 95% CI). B) For the top 5,000 abundant genes, the log2 fold-change in relative abundances between the two MEM-treated biopsies and the two control biopsies were plotted. C) Sampling graphic for collection of 12 biopsies from each of 5 participants (3 biopsies taken from 4 separate regions of the GI tract). Biopsies within one region were sampled within one field of view (5-cm diameter).

Data Availability

The datasets generated during and analyzed during the current study will be made available upon publication at CaltechDATA, URL pending. Microbial sequencing data will be made available upon publication at NCBI Accession doi pending.

Code Availability

The code utilized in data processing and analysis will be made available upon publication at CaltechDATA, URL pending.

It would be great to share the code for the review process as well.

We are happy to share the code with the Reviewers. We have submitted a zipped file “MEM_Submission.zip” containing all the code and data utilized in this manuscript.

All data and code will be made freely available on CaltechDATA at the time of publication.

Decision Letter, second revision:

Our ref: NMETH-A50720B

13th Jun 2023

Dear Dr. Ismagilov,

Thank you for submitting your revised manuscript "Microbial-enrichment method enables high-throughput metagenomic characterization from host-rich samples" (NMETH-A50720B). It has now been seen by the original referees and their comments are below. The reviewers find that the paper has improved in revision, and therefore we'll be happy in principle to publish it in Nature Methods, pending minor revisions to satisfy the referees' final requests and to comply with our editorial and formatting guidelines.

We are now performing detailed checks on your paper and will send you a checklist detailing our editorial and formatting requirements in about a week. Please do not upload the final materials and make any revisions until you receive this additional information from us.

TRANSPARENT PEER REVIEW

Nature Methods offers a transparent peer review option for new original research manuscripts submitted from 17th February 2021. We encourage increased transparency in peer review by publishing the reviewer comments, author rebuttal letters and editorial decision letters if the authors agree. Such peer review material is made available as a supplementary peer review file. **Please state in the cover letter 'I wish to participate in transparent peer review' if you want to opt in, or 'I do not wish to participate in transparent peer review' if you don't.** Failure to state your preference will result in delays in accepting your manuscript for publication.

Please note: we allow redactions to authors' rebuttal and reviewer comments in the interest of confidentiality. If you are concerned about the release of confidential data, please let us know specifically what information you would like to have removed. Please note that we cannot incorporate redactions for any other reasons. Reviewer names will be published in the peer review files if the reviewer signed the comments to authors, or if reviewers explicitly agree to release their name. For more information, please refer to our [FAQ page](https://www.nature.com/documents/nr-transparent-peer-review.pdf).

ORCID

IMPORTANT: Non-corresponding authors do not have to link their ORCIDs but are encouraged to do so. Please note that it will not be possible to add/modify ORCIDs at proof. Thus, please let your co-authors know that if they wish to have their ORCID added to the paper they must follow the procedure described in the following link prior to acceptance:

<https://www.springernature.com/gp/researchers/orcid/orcid-for-nature-research>

Thank you again for your interest in Nature Methods. Please do not hesitate to contact me if you have any questions. We will be in touch again soon.

Sincerely,
Lei

Lei Tang, Ph.D.
Senior Editor
Nature Methods

Reviewer #1 (Remarks to the Author):

This revised manuscript comprehensively addresses the reviews (I agree that comparisons to mock communities, which have their own problems, are not needed given the data provided, and also that for clinical cases it is not surprising that the stool samples are this different from one another as subject to subject variation is often high in such situations). The technique will be widely useful in the community. The authors may want to consider whether they will submit their data to SRA or if it needs to go to dbGAP given recent news about the failure of computational host filtering to adequately remove human DNA sequences.

Reviewer #2 (Remarks to the Author):

In general, the authors have addressed my major concerns.
There are still a few concerns need to be clarified and addressed:

Initial response 2. Commercial kits can achieve good DNA extraction efficiency in tissue samples with some modifications, such as trypsin digestion prior to QIAamp kit (Science. 2016 Nov 11;354(6313):744-747). This should also be discussed.

The author have added this citation to the manuscript: To prevent the majority of shotgun-sequencing reads assigning to host, a wide variety of host-removal (i.e. host- depletion) methods have been developed²¹⁻²⁶; however, commercial and published protocols are primarily used on liquid samples, and none have been shown to be sufficiently effective to enable bacterial MAG construction from solid mammalian tissues.

Although the article (Science. 2016 Nov 11;354(6313):744-747) did not evaluate the effectiveness of host DNA depletion, I believe the author should specify this method and its potential for success in host DNA depletion in tissue sample.

Besides, the authors argued that trypsin digestion prior to QIAamp kit might not be a good method for tissue sample host DNA depletion, "we expect Proteinase K to be similar if not more effective than a trypsin digestion because Proteinase K is known to produce shorter peptides than trypsin (see Dau et al. 2020; PMC7377536). Given that the Proteinase K pretreatment followed by saponin-selective lysis did not produce comparable host removal to MEM, we would expect the same to be true of trypsin digestion prior to QIAamp." Although the author chose not to get into these speculative discussions in the revised manuscript, I respectfully disagree with the authors' viewpoint. The reason of employing trypsin digestion prior to using the QIAamp Microbiome Kit is to enzymatically digest the tissue into single cells, enabling the sample to become more compatible with the protocol for the QIAamp Microbiome Kit. In this regard, I do not believe that Proteinase K and trypsin are directly comparable.

Initial response 3: It is impressive that MEM has superior performance when compared with two commercial kits and lyPMA. I think the authors should also consider comparing MEM to a highly cited method using 2.2% saponin, which allows up to 99.99% of host nucleic acid removal in clinical samples (Nat Biotechnol. 2019 Jul;37(7):783-792). Although saponin is present in AHL buffer of QIAamp, its concentration (7.73 wt-%) is much higher.

The authors have updated a statement in the results, "By comparing paired host-depleted and control samples, we found that lyPMA and QIAamp induced the most total bacterial losses whereas MolYsis and QIAamp induced the least uniform bacterial losses, with some taxa dropping more than 100-fold (Fig 1C-D). Previous literature suggests QIAamp's saponin concentration can be lowered to limit some of these bacterial losses."

I don't think 2.2% saponin is directly comparable to the AHL buffer in the QIAamp kit. In addition to higher concentrations of saponin, AHL also contains other components such as polyanethol sulfonic acid sodium salt. I respectfully disagree with the authors' viewpoint: "We would expect a higher concentration of saponin to be more effective at lysing host cells at the expense of inducing further microbial bias. Because we have tested the higher concentration (7.73 wt%) of saponin already (through the QIAamp kit), we would not predict a lower concentration (2.2 wt%) to improve host removal, although it may reduce bacterial losses induced by QIAamp." The choice of saponin concentration involves a trade-off between the effectiveness of cell lysis and its impact on microbial

lysis. In the study by Charalampous et al. (Nat Biotechnol. 2019 Jul;37(7):783-792), they demonstrated that a 2.2% saponin-based method for host DNA depletion removes up to 99.99% of host nucleic acids from clinical respiratory samples and. They also tested a concentration of 2.5% saponin but ultimately chose 2.2% as the optimal concentration.

Since the author did not directly compare MEM with some previous good approaches (such as Nat Biotechnol. 2019 Jul;37(7):783-792, Science. 2016 Nov 11;354(6313):744-747) in experiments, I suggest the author should mention these good approaches as objectively as possible in the background section of the study, rather than vaguely describing them.

11. Contaminant DNA has a negative impact on microbiome research, particularly when evaluating samples with low microbial biomass. It can come from a variety of sources, including laboratory surroundings, researchers, and laboratory reagents. The authors used at least three processing blanks for each DNA extraction batch as a baseline for host DNA detection, but did not set blank controls at the host-DNA depletion step to monitor microbial contamination from the 1.4mm beads-beating procedure. To ensure that the increased taxon (compared to non-host-depletion groups) identified in MEM-treated samples is not the result of contamination, blank controls for MEM should be established.

According to the author, "for the intestinal biopsies, the bacterial load ranged dramatically (see Fig S5), with some samples having similar bacterial load to the processing blanks." However, Figure S5 shows the bacterial load in saliva and extraction/processing blanks but not bacterial load in biopsies.

Finally, I have one more suggestion. Regarding the authors' method of quantifying human DNA using single-copy gene primers, it is important to consider that the human genome is approximately 1000 times larger than that of bacteria. Even detecting just one copy of a human gene would still indicate a significant amount of human DNA background. Therefore, I believe that the most crucial metric is not the fold reduction of host DNA measured by ddPCR but rather the proportion of microbial data compared to the total data after sequencing. The more sequencing data generated, the lower the proportion of human DNA, allowing for a higher representation of microbial data. This is especially critical for subsequent genome assembly and the discovery of microbial species diversity. Although Figure 5A showed that ~10% of reads were identified as non-host in MEM-treated samples whereas less than 0.01% of reads were identified as non-host in the untreated controls, the sample size was small and all from one individual. Due to the varying levels (for the intestinal biopsies, the bacterial load ranged dramatically with some samples having similar bacterial load to the processing blank) in bacterial load in colon biopsies, I suggest the authors discuss whether the quantity of bacteria has an impact on the efficiency of MEM treatment. Specifically, it would be valuable to investigate whether non-host reads can still reach 10% after MEM treatment when the bacterial load is extremely low. Additionally, it is hoped that the authors can extend the discussion to evaluate the suitability of MEM for tissues with even lower bacterial loads, such as gastric mucosa.

Reviewer #3 (Remarks to the Author):

authors have responded to most of my concerns

Author Rebuttal, second revision:

Point-By-Point Response to Reviewers

(New additions to the manuscript and SI are indicated in red text)

Reviewer #1 (Remarks to the Author):

This revised manuscript comprehensively addresses the reviews (I agree that comparisons to mock communities, which have their own problems, are not needed given the data provided, and also that for clinical cases it is not surprising that the stool samples are this different from one another as subject to subject variation is often high in such situations). The technique will be widely useful in the community. The authors may want to consider whether they will submit their data to SRA or if it needs to go to dbGAP given recent news about the failure of computational host filtering to adequately remove human DNA sequences.

We thank the Reviewer for all of their comments during this process. We also appreciate and share this concern about inadvertently publishing host reads. We have submitted this project to SRA and we are utilizing their HRRT pipeline to remove the host reads ([Scrubbing human sequence contamination from Sequence Read Archive \(SRA\) submissions - NCBI Insights \(nih.gov\)](#)). As an additional sanity check, we will also check HRRT host removal performance by confirming identity of remaining reads are non-human derived.

Reviewer #2 (Remarks to the Author):

In general, the authors have addressed my major concerns.
There are still a few concerns need to be clarified and addressed:

Initial response 2. Commercial kits can achieve good DNA extraction efficiency in tissue samples with some modifications, such as trypsin digestion prior to QIAamp kit (Science. 2016 Nov 11;354(6313):744-747). This should also be discussed.

The author have added this citation to the manuscript: To prevent the majority of shotgun-sequencing reads assigning to host, a wide variety of host-removal (i.e. host- depletion) methods have been developed²¹⁻²⁶; however, commercial and published protocols are primarily used on liquid samples, and none have been shown to be sufficiently effective to enable bacterial MAG

construction from solid mammalian tissues.

Although the article (Science. 2016 Nov 11;354(6313):744-747) did not evaluate the effectiveness of host DNA depletion, I believe the author should specify this method and its potential for success in host DNA depletion in tissue sample.

Besides, the authors argued that trypsin digestion prior to QIAamp kit might not be a good method for tissue sample host DNA depletion, “we expect Proteinase K to be similar if not more effective than a trypsin digestion because Proteinase K is known to produce shorter peptides than trypsin (see Dau et al. 2020; PMC7377536). Given that the Proteinase K pretreatment followed by saponin-selective lysis did not produce comparable host removal to MEM, we would expect the same to be true of trypsin digestion prior to QIAamp.” Although the author chose not to get into these speculative discussions in the revised manuscript, I respectfully disagree with the authors' viewpoint. The reason of employing trypsin digestion prior to using the QIAamp Microbiome Kit is to enzymatically digest the tissue into single cells, enabling the sample to become more compatible with the protocol for the QIAamp Microbiome Kit. In this regard, I do not believe that Proteinase K and trypsin are directly comparable.

We thank the Reviewer for their thoughtful insight into this protocol. In order to avoid any speculative discussion, we have added the following statement to the Introduction, citing the important work that the Reviewer has identified (we note that Science. 2016 Nov 11;354(6313):744-747 is reference 24) and also emphasizing these protocols have the potential to be successful in these sample types.

To prevent the majority of shotgun-sequencing reads assigning to host, a wide variety of host-removal (i.e. host-depletion) methods have been developed.²¹⁻²⁶ Published and commercial good protocols have enabled both long-read sequencing and bacterial MAG construction from mammalian derived liquid samples. Although some protocols have also been validated for use on solid tissue sample types, and others may have potential for success in these samples, none have been shown to be sufficiently effective to enable bacterial MAG construction from solid mammalian tissues. Additionally, many host-depletion methods are not feasible to perform in the clinic due to extensive processing times and complex protocols.

Initial response 3: It is impressive that MEM has superior performance when compared with two commercial kits and lyPMA. I think the authors should also consider comparing MEM to a highly cited method using 2.2% saponin, which allows up to 99.99% of host nucleic acid removal in clinical samples (Nat Biotechnol. 2019 Jul;37(7):783-792). Although saponin is present in AHL buffer of QIAamp, its concentration (7.73 wt-%) is much higher.

The authors have updated a statement in the results, “By comparing paired host-depleted and control samples, we found that lyPMA and QIAamp induced the most total bacterial losses whereas MoLYsis and QIAamp induced the least uniform bacterial losses, with some taxa dropping more than 100-fold (Fig 1C-D). Previous literature suggests QIAamp’s saponin concentration can be lowered to limit some of these bacterial losses.”

I don't think 2.2% saponin is directly comparable to the AHL buffer in the QIAamp kit. In addition to higher concentrations of saponin, AHL also contains other components such as polyanethol sulfonic acid sodium salt. I respectfully disagree with the authors' viewpoint: "We would expect a higher concentration of saponin to be more effective at lysing host cells at the expense of inducing further microbial bias. Because we have tested the higher concentration (7.73 wt%) of saponin already (through the QIAamp kit), we would not predict a lower concentration (2.2 wt%) to improve host removal, although it may reduce bacterial losses induced by QIAamp." The choice of saponin concentration involves a trade-off between the effectiveness of cell lysis and its impact on microbial lysis. In the study by Charalampous et al. (Nat Biotechnol. 2019 Jul;37(7):783-792), they demonstrated that a 2.2% saponin-based method for host DNA depletion removes up to 99.99% of host nucleic acids from clinical respiratory samples and. They also tested a concentration of 2.5% saponin but ultimately chose 2.2% as the optimal concentration.

Since the author did not directly compare MEM with some previous good approaches (such as Nat Biotechnol. 2019 Jul;37(7):783-792, Science. 2016 Nov 11;354(6313):744-747) in experiments, I suggest the author should mention these good approaches as objectively as possible in the background section of the study, rather than vaguely describing them.

We thank the Reviewer for their thoughtful insight into this protocol. We agree these are good protocols with potential for success. We have added the following statement to the Introduction section (we note that Nat Biotechnol. 2019 Jul;37(7):783-792 is reference 25):

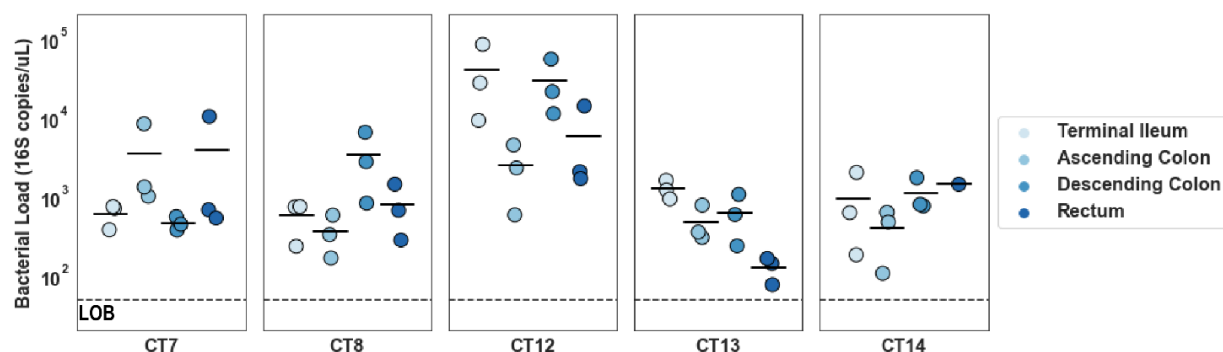
To prevent the majority of shotgun-sequencing reads assigning to host, a wide variety of host-removal (i.e. host-depletion) methods have been developed.²¹⁻²⁶ Published and commercial good protocols have enabled both long-read sequencing and bacterial MAG construction from mammalian derived liquid samples. Although some protocols have also been validated for use on solid tissue sample types, and others may have potential for success in these samples, none have been shown to be sufficiently effective to enable bacterial MAG construction from solid mammalian tissues. Additionally, many host-depletion methods are not feasible to perform in the clinic due to extensive processing times and complex protocols.

11. Contaminant DNA has a negative impact on microbiome research, particularly when evaluating

samples with low microbial biomass. It can come from a variety of sources, including laboratory surroundings, researchers, and laboratory reagents. The authors used at least three processing blanks for each DNA extraction batch as a baseline for host DNA detection, but did not set blank controls at the host-DNA depletion step to monitor microbial contamination from the 1.4mm beads-beating procedure. To ensure that the increased taxon (compared to non-host-depletion groups) identified in MEM-treated samples is not the result of contamination, blank controls for MEM should be established.

According to the author, "for the intestinal biopsies, the bacterial load ranged dramatically (see Fig S5), with some samples having similar bacterial load to the processing blanks." However, Figure S5 shows the bacterial load in saliva and extraction/processing blanks but not bacterial load in biopsies.

We thank the Reviewer for catching this typographical error. We have previously addressed this issue in the paper itself by including Fig S5 and S7 (currently Extended Data Fig 2). However, in our response to the Reviewer comments, we should have included references to both Fig S5 and Extended Data Fig 2 (not just Fig S5 as we originally did). We have updated Extended Data Fig2 to be in units of 16S rRNA gene Copies per uL of elution, so the load can be compared to the processing blanks.



Extended Data Figure 2: **Bacterial loads of longitudinal biopsies.** 16S rRNA gene copies were quantified as a proxy for bacterial load for all biopsies. Samples were plotted by participant and then by location. (N=3 biological replicates for each location for each patient, **LOB refers to limit of blank defined as $\text{LoB} = \text{mean}_{\text{blank}} + 1.645[\text{SD}_{\text{blank}}]$ based on three processing blanks.**)

Finally, I have one more suggestion. Regarding the authors' method of quantifying human DNA using single-copy gene primers, it is important to consider that the human genome is approximately 1000 times larger than that of bacteria. Even detecting just one copy of a human gene would still indicate a

significant amount of human DNA background. Therefore, I believe that the most crucial metric is not the fold reduction of host DNA measured by ddPCR but rather the proportion of microbial data compared to the total data after sequencing. The more sequencing data generated, the lower the proportion of human DNA, allowing for a higher representation of microbial data. This is especially critical for subsequent genome assembly and the discovery of microbial species diversity. Although Figure 5A showed that ~10% of reads were identified as non-host in MEM-treated samples whereas less than 0.01% of reads were identified as non-host in the untreated controls, the sample size was small and all from one individual. Due to the varying levels (for the intestinal biopsies, the bacterial load ranged dramatically with some samples having similar bacterial load to the processing blank) in bacterial load in colon biopsies, I suggest the authors discuss whether the quantity of bacteria has an impact on the efficiency of MEM treatment. Specifically, it would be valuable to investigate whether non-host reads can still reach 10% after MEM treatment when the bacterial load is extremely low. Additionally, it is hoped that the authors can extend the discussion to evaluate the suitability of MEM for tissues with even lower bacterial loads, such as gastric mucosa.

We thank the Reviewer for bringing up this great point. We agree a quantification of the percentage of non-host reads is more important for users than quantification of host removal from ddPCR measurements. As Reviewer is suggesting, and as we have shown in Extended Data Figure 1, the percentage of non-host reads is dependent on bacterial load. While the efficiency of MEM treatment is fairly constant (>2000 fold removal), the resulting ratio of bacterial and human reads depends on the starting bacterial load. This leads to a wide distribution in the percentage of non-host reads after sequencing, ranging from 0.01% to 89%. As a result, we utilized the ddPCR measurements to quantify both bacterial load and host removal, so potential users can refer to Figure Extended Data Figure 1 to estimate what percentage of non-host reads they may expect based on quantification of bacterial load. We have added a sentence in the Discussion section reflecting this:

We acknowledge the following limitations of the MEM approach demonstrated here. We have analyzed biopsies with as few as 10^2 16S rRNA gene copies/ μ L in the 100 μ L elution from the extraction column (corresponding to $\sim 10^4$ 16S copies/mg of tissue), however deep analysis of samples below this bacterial load will require greater levels of host depletion and/or greater sequencing depth (Fig ED1-2). **We advise users to refer to Fig ED1 to predict the percentage of non-host reads from bacterial load to guide sequencing depth decisions.**

Reviewer #3 (Remarks to the Author):

authors have responded to most of my concerns

We thank the Reviewer for all their assistance and insight during this process.

Final Decision Letter:

27th Aug 2023

Dear Professor Ismagilov,

I am pleased to inform you that your Article, "Microbial-enrichment method enables high-throughput metagenomic characterization from host-rich samples", has now been accepted for publication in Nature Methods. Your paper is tentatively scheduled for publication in our November print issue, and will be published online prior to that. The received and accepted dates will be 29th Nov 2022 and 27th Aug 2023. This note is intended to let you know what to expect from us over the next month or so, and to let you know where to address any further questions.

Acceptance is conditional on the data in the manuscript not being published elsewhere, or announced in the print or electronic media, until the embargo/publication date. These restrictions are not intended to deter you from presenting your data at academic meetings and conferences, but any enquiries from the media about papers not yet scheduled for publication should be referred to us.

Over the next few weeks, your paper will be copyedited to ensure that it conforms to Nature Methods style. Once your paper is typeset, you will receive an email with a link to choose the appropriate publishing options for your paper and our Author Services team will be in touch regarding any additional information that may be required.

You will receive a link to your electronic proof via email with a request to make any corrections within 48 hours. If, when you receive your proof, you cannot meet this deadline, please inform us at rjsproduction@springernature.com immediately.

Please note that *Nature Methods* is a Transformative Journal (TJ). Authors may publish their research with us through the traditional subscription access route or make their paper immediately open access through payment of an article-processing charge (APC). Authors will not be required to make a final decision about access to their article until it has been accepted. [Find out more about Transformative Journals](https://www.springernature.com/gp/open-research/transformative-journals)

Authors may need to take specific actions to achieve [compliance](https://www.springernature.com/gp/open-research/funding/policy-compliance-faqs) with funder and institutional open access mandates. If your research is supported by a funder that requires immediate open access (e.g. according to [Plan S principles](https://www.springernature.com/gp/open-research/plan-s-compliance)) then you should select the gold OA route, and we will direct you to the compliant route where possible. For authors selecting the subscription publication route, the journal's standard licensing terms will need to be accepted, including [66](https://www.springernature.com/gp/open-</p>
</div>
<div data-bbox=)

research/policies/journal-policies">self-archiving policies. Those licensing terms will supersede any other terms that the author or any third party may assert apply to any version of the manuscript.

If you have any questions about our publishing options, costs, Open Access requirements, or our legal forms, please contact ASJournals@springernature.com

Your paper will now be copyedited to ensure that it conforms to Nature Methods style. Once proofs are generated, they will be sent to you electronically and you will be asked to send a corrected version within 24 hours. It is extremely important that you let us know now whether you will be difficult to contact over the next month. If this is the case, we ask that you send us the contact information (email, phone and fax) of someone who will be able to check the proofs and deal with any last-minute problems.

If, when you receive your proof, you cannot meet the deadline, please inform us at rjsproduction@springernature.com immediately.

Once your manuscript is typeset and you have completed the appropriate grant of rights, you will receive a link to your electronic proof via email with a request to make any corrections within 48 hours. If, when you receive your proof, you cannot meet this deadline, please inform us at rjsproduction@springernature.com immediately.

Once your paper has been scheduled for online publication, the Nature press office will be in touch to confirm the details.

If you have posted a preprint on any preprint server, please ensure that the preprint details are updated with a publication reference, including the DOI and a URL to the published version of the article on the journal website.

Once your paper has been scheduled for online publication, the Nature press office will be in touch to confirm the details.

Content is published online weekly on Mondays and Thursdays, and the embargo is set at 16:00 London time (GMT)/11:00 am US Eastern time (EST) on the day of publication. If you need to know the exact publication date or when the news embargo will be lifted, please contact our press office after you have submitted your proof corrections. Now is the time to inform your Public Relations or Press Office about your paper, as they might be interested in promoting its publication. This will allow them time to prepare an accurate and satisfactory press release. Include your manuscript tracking number NMETH-A50720C and the name of the journal, which they will need when they contact our office.

About one week before your paper is published online, we shall be distributing a press release to news organizations worldwide, which may include details of your work. We are happy for your institution or funding agency to prepare its own press release, but it must mention the embargo date and Nature Methods. Our Press Office will contact you closer to the time of publication, but if you or your Press Office have any inquiries in the meantime, please contact press@nature.com.

To assist our authors in disseminating their research to the broader community, our SharedIt initiative provides you with a unique shareable link that will allow anyone (with or without a subscription) to

read the published article. Recipients of the link with a subscription will also be able to download and print the PDF.

As soon as your article is published, you will receive an automated email with your shareable link.

You can now use a single sign-on for all your accounts, view the status of all your manuscript submissions and reviews, access usage statistics for your published articles and download a record of your refereeing activity for the Nature journals.

Nature Portfolio journals [encourage authors to share their step-by-step experimental protocols](https://www.nature.com/nature-research/editorial-policies/reporting-standards#protocols) on a protocol sharing platform of their choice. Nature Portfolio's Protocol Exchange is a free-to-use and open resource for protocols; protocols deposited in Protocol Exchange are citable and can be linked from the published article. More details can found at www.nature.com/protocolexchange/about.

Please note that you and any of your coauthors will be able to order reprints and single copies of the issue containing your article through Nature Portfolio's reprint website, which is located at <http://www.nature.com/reprints/author-reprints.html>. If there are any questions about reprints please send an email to author-reprints@nature.com and someone will assist you.

Please feel free to contact me if you have questions about any of these points.

Best regards,
Lei

Lei Tang, Ph.D.
Senior Editor
Nature Methods