Ecology
Computational and Systems Biology

Microbes with higher metabolic independence are enriched in human gut microbiomes under stress

Reviewed Preprint
v2 • November 20, 2024
Revised by authors

Reviewed Preprint
v1 • September 8, 2023

Iva Veseli, Yiqun T Chen, Matthew S Schechter, Chiara Vanni, Emily C Fogarty, Andrea R Watson, Bana Jabri, Ran Blekhman, Amy D Willis, Michael K Yu, Antonio Fernàndez-Guerra, Jessika Füssel, A Murat Eren

Biophysical Sciences Program, The University of Chicago, Chicago, USA • Department of Medicine, The University of Chicago, Chicago, USA • Data Science Institute and Department of Biomedical Data Science, Stanford University, Stanford, USA • Committee on Microbiology, The University of Chicago, Chicago, USA • MARUM Center for Marine Environmental Sciences, University of Bremen, Bremen, Germany • Department of Biostatistics, University of Washington, Seattle, USA • Toyota Technological Institute at Chicago, Chicago, USA • Lundbeck Foundation GeoGenetics Centre, GLOBE Institute, University of Copenhagen, Copenhagen, Denmark • Institute for Chemistry and Biology of the Marine Environment, University of Oldenburg, Oldenburg, Germany • Marine 'Omics Bridging Group, Max Planck Institute for Marine Microbiology, Bremen, Germany • Alfred Wegener Institute for Polar and Marine Research, Bremerhaven, Germany • Helmholtz Institute for Functional Marine Biodiversity, Oldenburg, Germany

eLife Assessment

This study presents an **important** new bioinformatics tool for normalizing gene copy number from metagenomic assemblies and applies it to gain functional insights into the loss of microbial diversity during conditions of stress. The inclusion of extensive computational validation makes this a **compelling** study that raises intriguing new hypotheses regarding the impact of disease states on the gut microbiome. This paper will likely be of broad interest to researchers studying the role of complex microbial communities in host health and disease.

https://doi.org/10.7554/eLife.89862.2.sa4

Abstract

A wide variety of human diseases are associated with loss of microbial diversity in the human gut, inspiring a great interest in the diagnostic or therapeutic potential of the microbiota. However, the ecological forces that drive diversity reduction in disease states remain unclear, rendering it difficult to ascertain the role of the microbiota in disease emergence or severity. One hypothesis to explain this phenomenon is that microbial diversity is diminished as disease states select for microbial populations that are more fit to survive environmental stress caused by inflammation or other host factors. Here, we tested this hypothesis on a large scale, by developing a software framework to quantify the enrichment of microbial metabolisms in complex metagenomes as a function of microbial diversity. We

applied this framework to over 400 gut metagenomes from individuals who are healthy or diagnosed with inflammatory bowel disease (IBD). We found that high metabolic independence (HMI) is a distinguishing characteristic of microbial communities associated with individuals diagnosed with IBD. A classifier we trained using the normalized copy numbers of 33 HMI-associated metabolic modules not only distinguished states of health versus IBD, but also tracked the recovery of the gut microbiome following antibiotic treatment, suggesting that HMI is a hallmark of microbial communities in stressed gut environments.

Introduction

The human gut is home to a diverse assemblage of microbial cells that form complex communities (Coyte, Schluter, and Foster 2015). This gut microbial ecosystem is established almost immediately after birth and plays a lifelong role in human wellbeing by contributing to immune system maturation and functioning (Belkaid and Hand 2014; Maynard et al. 2012), extracting dietary nutrients (Hijova 2019), providing protection against pathogens (Khosravi and Mazmanian 2013), metabolizing drugs (M. Zimmermann et al. 2019), and more (Knight et al. 2017). There is no universal definition of a healthy gut microbiome (Fan and Pedersen 2021), but associations between host disease states and changes in microbial community composition have sparked great interest in the therapeutic potential of gut microbes (Cani 2018; Sorbara and Pamer 2022) and led to the emergence of hypotheses that directly link disruptions of the gut microbiome to non-communicable diseases of complex etiology (Byndloss and Bäumler 2018).

Inflammatory bowel diseases (IBDs), which describe a heterogeneous group of chronic inflammatory disorders (Shan, Lee, and Chang 2022), represent an increasingly common health risk around the globe (Kaplan 2015). Understanding the role of gut microbiota in IBD has been a major area of focus in human microbiome research. Studies focusing on individual microbial taxa that typically change in relative abundance in IBD patients have proposed a range of host-microbe interactions that may contribute to disease manifestation and progression (Joossens et al. 2011; Schirmer et al. 2019; Henke et al. 2019; Machiels et al. 2014). However, even within well-constrained cohorts, a large proportion of variability in the taxonomic composition of the microbiota is unexplained, and the proportion of variability explained by disease status is low (Gevers et al. 2014; Schirmer et al. 2018; Lloyd-Price et al. 2019; Khan et al. 2019). As neither individual taxa nor broad changes in microbial community composition yield effective predictors of disease (Knox et al. 2019; M. Lee and Chang 2021), the role of gut microbes in the etiology of IBD – or the extent to which they are bystanders to disease – remains unclear (Khan et al. 2019).

The marked decrease in microbial diversity in IBD is often associated with the loss of Firmicutes populations and an increased representation of a relatively small number of taxa, such as *Bacteroides*, *Enterococcaceae*, and others (Prindiville et al. 2000; Saitoh et al. 2002; Sartor 2006; Rhodes 2007; Devkota et al. 2012; Machiels et al. 2014; Vineis et al. 2016; Lloyd-Price et al. 2019). Why a handful of taxa that also typically occur in healthy individuals in lower abundances (M. Lee and Chang 2021; Nishida et al. 2018) tend to dominate the IBD microbiome is a fundamental but open question to gain insights into the ecological underpinnings of the gut microbial ecosystem under IBD. Going beyond taxonomic summaries, a recent metagenome-wide metabolic modeling study revealed a significant loss of cross-feeding partners as a hallmark of IBD, where microbial interactions were disrupted in IBD-associated microbial communities compared to those found in healthy individuals (Marcelino et al. 2023). This observation is in line with another recent work that proposed that the extent of 'metabolic independence' (characterized by the genomic presence of a set of key metabolic modules for the synthesis of essential nutrients) is a determinant of microbial survival in IBD (Watson et al. 2023). It is conceivable that the disrupted metabolic interactions among microbes observed in

IBD (Marcelino et al. 2023) indicates an environment that lacks the ecosystem services provided by a complex network of microbial interactions, and selects for those organisms that harness high metabolic independence (HMI) (Watson et al. 2023). This interpretation offers an ecological mechanism to explain the dominance of populations with specific metabolic features in IBD. However, this proposed mechanism warrants further investigation.

Here we implemented a high-throughput, taxonomy-independent strategy to estimate metabolic capabilities of microbial communities directly from metagenomes and investigate whether the enrichment of populations with high metabolic independence predicts IBD in the human gut. We benchmarked our findings using representative genomes associated with the human gut and their distribution in healthy individuals and those who have been diagnosed with IBD. Our results suggest that high metabolic potential (indicated by a set of 33 largely biosynthetic metabolic pathways) provides enough signal to consistently distinguish gut microbiomes under stress from those that are in homeostasis, providing deeper insights into adaptive processes initiated by stress conditions that promote rare members of gut microbiota to dominance during disease.

Results and Discussion

We compiled 2,893 publicly-available stool metagenomes from 13 different studies, 5 of which explicitly studied the IBD gut microbiome (Supplementary Table 1a-c). The average sequencing depth varied across individual datasets (4.2 Mbp to 60.3 Mbp, with a median value of 21.4 Mbp, Supplementary Table 1c). To improve the sensitivity and accuracy of our downstream analyses that depend on metagenomic assembly, we excluded samples with less than 25 million reads, resulting in a set of 408 relatively deeply-sequenced metagenomes from 10 studies (26.4 Mbp to 61.9 Mbp, with a median value of 37.0 Mbp, Supplementary Table 1b, Supplementary File 1, Methods), which we *de novo* assembled individually. The final dataset included individuals who were healthy (n=229), diagnosed with IBD (n=101), or suffered from other gastrointestinal conditions ("non-IBD", n=78). In accordance with previous observations of reduced microbial diversity in IBD (Kostic, Xavier, and Gevers 2014; Nagalingam and Lynch 2012; Knox et al. 2019), the estimated number of populations based on the occurrence of bacterial single-copy core genes present in these metagenomes was higher in healthy individuals than those diagnosed with IBD (**Supplementary Figure 1**, Supplementary Table 1).

Estimating normalized copy numbers of metabolic pathways from metagenomic assemblies

Gaining insights into microbial metabolism requires accurate estimates of pathway presence/absence and completion. While a myriad of tools address this task for single genomes (Machado et al. 2018; Aziz et al. 2008; Arkin et al. 2018; Palù et al. 2022; Shaffer et al. 2020; Geller-McGrath et al. 2023; Zorrilla et al. 2021; Zhou et al., n.d.; J. Zimmermann, Kaleta, and Waschina 2021), working with complex environmental metagenomes poses additional challenges due to the large number of organisms that are present in metagenomic assemblies. A few tools can estimate community-level metabolic potential from metagenomes without relying on the reconstruction of individual population genomes or reference-based approaches (Ye and Doak 2009; Karp et al. 2021) (Supplementary Table 5). These high-level summaries of pathway presence and redundancy in a given environment are suitable for most surveys of metabolic capacity, particularly for microbial communities of similar richness. However, since the frequency of observed metabolic modules will increase as the number of distinct microbial populations in a habitat increases, investigations of metabolic determinants of survival across environmental conditions with substantial differences in microbial richness may suffer from ambiguous observations from quantitative data. For instance, the estimated copy number of a given metabolic module may be identical between two metagenomes but its enrichment may be relatively higher in the metagenome with a lower alpha diversity, revealing its potential role in overcoming

environment-specific selective pressures that influence an entire community. Working solely with raw copy numbers of metabolic modules without a normalization step that considers the microbial richness will thus shroud potentially critical insights. To quantify the differential enrichment of metabolic modules between metagenomes generated from healthy individuals and those from individuals diagnosed with IBD, we implemented a new software framework (*https://anvio.org/m/anvi-estimate-metabolism*) that reconstructs metabolic modules from genomes and metagenomes and then calculates the per-population copy number (PPCN) of modules in metagenomes to account for potential differences in microbial richness (Methods, Supplementary File 1 and 2). Briefly, the PPCN estimates the proportion of microbes in a community with a particular metabolic capacity (**Figure 1**, **Supplementary Figure 2**) by estimating the number of microbial populations in a given sample using single-copy core genes (SCGs) without having to reconstruct individual genomes first, thus maximizing the *de novo* recovery of gene content. Our validation of this method using simulated metagenomic data demonstrated that it is accurate in capturing metagenome-level metabolic capacity relative to genome-level metabolic capacity estimated from the same data (Supplementary File 2, Supplementary Table 6).

Key biosynthetic pathways are enriched in microbial populations from IBD samples

To gain insight into potential metabolic determinants of microbial survival in the IBD gut environment, we assessed the distribution of metabolic modules within samples from each group (IBD and healthy) with and without using PPCN normalization. A set of 33 metabolic modules were significantly enriched in metagenomes obtained from individuals diagnosed with IBD when PPCN normalization was applied (**Figure 2d, 2e**). Each metabolic module had an FDR-adjusted $p < 2e\text{-}10$ and an effect size > 0.12 from a Wilcoxon Rank Sum Test comparing IBD and healthy samples. The set included 17 modules that were previously associated with high metabolic independence (Watson et al. 2023) (**Figure 2f**). However, without PPCN normalization, the signal was masked by the overall higher copy numbers in healthy samples, and the same analysis did not detect higher metabolic potential in microbial populations associated with individuals diagnosed with IBD (**Figure 2a**), showing weaker differential occurrence between cohorts (**Figure 2b, 2c**, **Supplementary Figure 3**). This result suggests that the PPCN normalization is an important step in comparative analyses of metabolisms between samples with disparate levels of diversity.

The majority of the metabolic modules that were enriched in the microbiomes of IBD patients encoded biosynthetic capabilities (23 out of 33) that resolved to amino acid metabolism (33%), carbohydrate metabolism (21%), cofactor and vitamin biosynthesis (15%), nucleotide biosynthesis (12%), lipid biosynthesis (6%) and energy metabolism (6%) (Supplementary Table 2a). In contrast to previous reports based on reference genomes (Gevers et al. 2014; Morgan et al. 2012), amino acid synthesis and carbohydrate metabolism were not reduced in the IBD gut microbiome in our dataset. Rather, our results were in accordance with a more recent finding that predicted amino acid secretion potential is increased in the microbiomes of individuals with IBD (Heinken, Hertel, and Thiele 2021).

The metagenome-level enrichment of several key biosynthesis pathways supports the hypothesis that high metabolic independence (HMI) is a determinant of survival for microbial populations in the IBD gut environment. We investigated whether biosynthetic capacity in general was enriched in IBD samples, and 62 out of 88 (70%) biosynthesis pathways described in the KEGG database had a significant enrichment in the IBD sample group at an FDR-adjusted 5% significance level (**Supplementary Figure 5d**). However, a similar proportion of non-biosynthetic pathways, 63 out of 91 (69%), were also significantly increased in the IBD samples. While biosynthetic capacity

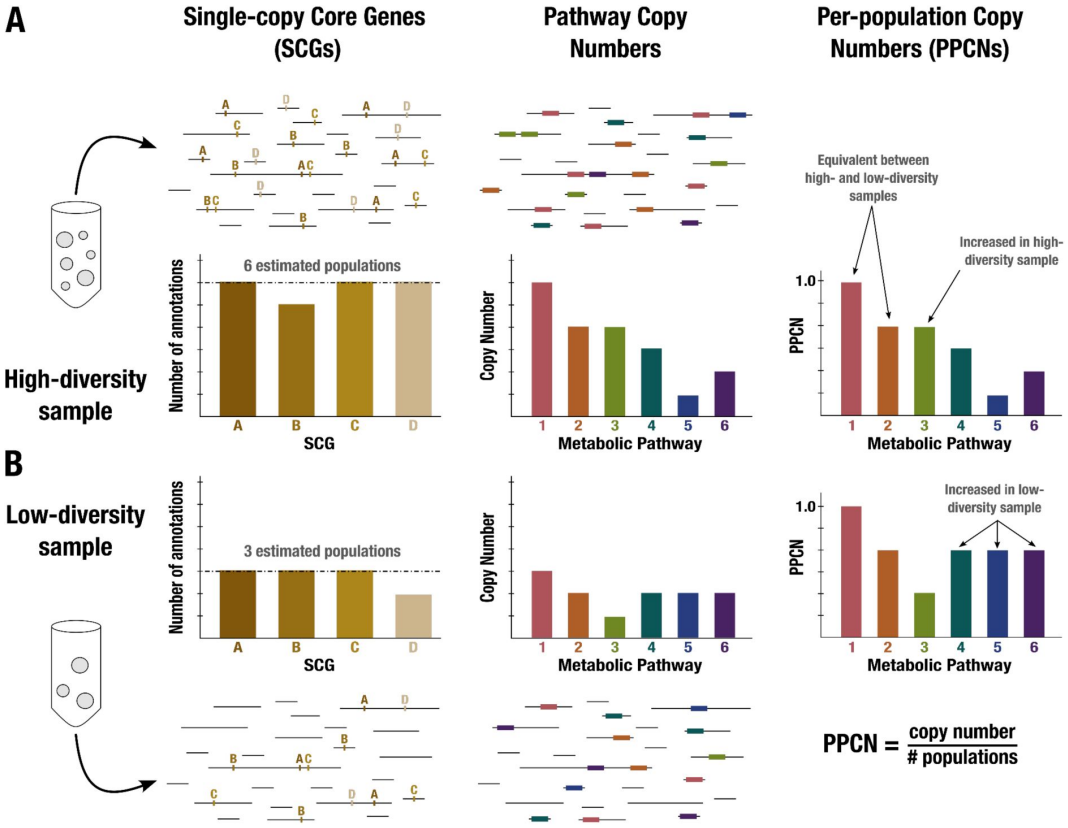


Figure 1.

Conceptual diagram of per-population copy number (PPCN) calculation.

Each step of the calculation is demonstrated in (**A**) for a sample with high diversity (6 microbial populations) and in (**B**) for a sample with low diversity (3 populations). Metagenome sequences are shown as black lines. The left panel shows the single-copy core genes annotated in the metagenome (indicated by letters), with a barplot showing the counts for different SCGs. The dashed black line indicates the mode of the counts, which is taken as the estimate of the number of populations. The middle panel shows the annotations of metabolic pathways (indicated by boxes and numerically labeled), with a barplot showing the copy number of each pathway (for more details on how this copy number is computed, see Supplementary File 1 and **Supplementary Figure 2**). The right panel shows the equation for per-population copy number (PPCN), with the barplots indicating the PPCN values for each metabolic pathway in each sample and arrows differentiating between different types of modules based on the comparison of their normalized copy numbers between samples.

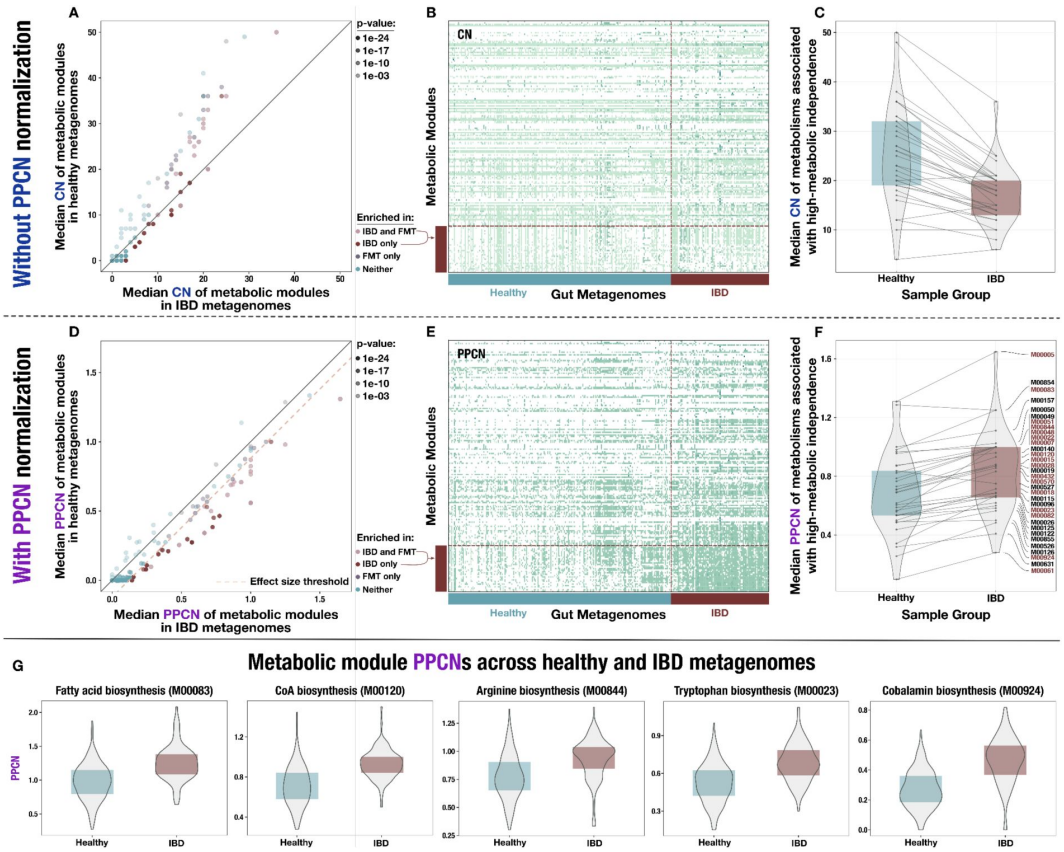


Figure 2.

Comparison of metabolic potential across healthy and IBD cohorts.

Panels **A – C** show unnormalized copy number data and the remaining panels show normalized per-population copy number (PPCN) data. **A)** Scatterplot of module copy number in IBD samples (x-axis) and healthy samples (y-axis). Transparency of points indicates the p-value of the module in a Wilcoxon Rank Sum test for enrichment (based on PPCN data), and color indicates whether the module is enriched in the IBD samples (in this study), enriched in the good colonizers from the fecal microbiota transplant (FMT) study (Watson et al. 2023), or enriched in both. **B)** Heatmap of unnormalized copy numbers for all modules. IBD-enriched modules are highlighted by the red bar on the left. Sample group is indicated by the blue (healthy) and red (IBD) bars on the bottom. **C)** Boxplots of median copy number for each module enriched in the FMT colonizers from (Watson et al. 2023) in the healthy samples (blue) and the IBD samples (red). Solid lines connect the same module in each plot. **D)** Scatterplot of module PPCN values in IBD samples (x-axis) and healthy samples (y-axis). Transparency and color of points are defined as in panel (A). The pink dashed line indicates the effect size threshold applied to modules when determining their enrichment in IBD. **E)** Heatmap of PPCN values for all modules. Side bars defined as in (B). **F)** Boxplots of median PPCN values for modules enriched in the FMT colonizers from (Watson et al. 2023) in the healthy samples (blue) and the IBD samples (red). Lines defined as in (D). Modules that were also enriched in the IBD samples (in this study) are highlighted in red. **G)** Boxplots of PPCN values for individual modules in the healthy samples (blue) and the IBD samples (red). All example modules were enriched in both this study and in (Watson et al. 2023).

is not over-represented in the IBD sample group compared to other types of metabolism, the high proportion of enriched pathways associated with biosynthesis suggests that biosynthetic capacity is important for microbial resilience.

Within our set of 33 pathways that were enriched in IBD, it is notable that all the biosynthesis and central carbohydrate pathways are directly or indirectly linked via shared enzymes and metabolites. Each enriched module shared on average 25.6% of its enzymes and 40.2% of metabolites with the other enriched modules, and overall 18.2% of enzymes and 20.4% of compounds across these pathways were shared (Supplementary Table 2a). Thus, modules may be enriched not just due to the importance of their immediate end products, but also because of their role in the larger metabolic network. The few standalone modules that were enriched included the efflux pump MepA and the beta-Lactam resistance system, which are associated with drug resistance. These capacities may provide an advantage since antibiotics are a common treatment for IBDs (Nitzan et al. 2016), but are not related to the systematic enrichment of biosynthesis pathways that likely provide resilience to general environmental stress rather than to a specific stressor such as antibiotics.

While so far we divided samples into two groups, our dataset also includes individuals who do not suffer from IBD, yet are not healthy either. A recent study using flux balance analysis to model metabolite secretion potential in the dysbiotic, non-dysbiotic, and control gut communities of Crohn's Disease patients has shown that several predicted microbial metabolic activities align with gradients of host health (Heinken, Hertel, and Thiele 2021). To test whether the HMI signal captures gradients in host health, we included the 'non-IBD' group of patients that suffer from gastrointestinal conditions other than IBD in our analysis. The set of 78 samples classified as 'non-IBD' indeed represent an intermediate group between healthy individuals and those diagnosed with IBD (**Supplementary Figure 5b**). While the HMI signal was reduced in 'non-IBD' patients, 75% of the pathways enriched in IBD patients were also enriched in the 'non-IBD' group compared to healthy individuals. Similarly, when sorting each individual cohort along a health gradient based on cohort descriptions in their respective studies (Supplementary File 1), the relative proportion of metabolic pathways indicative of HMI increased as a function of increasing disease severity (**Supplementary Figure 6a**). These findings suggest that the HMI signal is sufficiently sensitive to resolve gradients in host health and could serve as a diagnostic tool to monitor changing stress levels in a single individual over time.

Microbiome data generated by different groups can result in systematic biases that may outweigh biological differences between otherwise similar samples (Lozupone et al. 2013; Sinha et al. 2017; Clausen and Willis 2022). The potential impact of such biases constitutes an important consideration for meta-analyses such as ours that analyze publicly available metagenomes from multiple sources. To account for cohort biases, we conducted an analysis of our data on a per-cohort basis, which showed robust differences between the sample groups across multiple cohorts (**Supplementary Figure 6b, 6c**). Another source of potential bias in our results is due to the representation of microbial functions in genomes in publicly available databases. For instance, we noticed that, independent of the annotation strategy, a smaller proportion of genes resolved to known functions in metagenomic assemblies of the healthy samples compared to the assemblies we generated from the IBD group (**Supplementary Figure 4**). This highlights the possibility that healthy samples merely appear to harbor less metabolic capabilities due to missing annotations. Indeed, we found that the normalized copy numbers of most metabolic modules were reduced in the healthy group, where 84% of KEGG modules (98 out of 118) have significantly lower median copy numbers (**Supplementary Figure 5c**, Supplementary File 1). While the presence of a bias between the two cohorts is clear, the source of this bias and its implications are not as clear. One hypothesis that could explain this phenomenon is that the increased proportion of unknown functions in environments where populations with low metabolic independence (LMI) thrive is due to our inability to identify distant homologs of even well-studied functions in poorly studied novel genomes through public databases. If true, this would indeed impair our ability to annotate

genes using state-of-the-art functional databases, and bias metabolic module completion estimates. Such a limitation would indeed warrant a careful reconsideration of common workflows and studies that rely on public resources to characterize gene function in complex environments. Another hypothesis that could explain our observation is that the general absence in culture of microbes with smaller genomes (that likely fare better in diverse gut ecosystems) had a historical impact on the characterization of novel functions that represent a relatively larger fraction of their gene repertoire. If true, this would suggest that the unknown functions are unlikely essential for well-studied metabolic capabilities. Furthermore, HMI and LMI genomes may be indistinguishable with respect to the distribution of such novel genes, but the increased number of genes in HMI genomes that resolve to well-studied metabolisms would reduce the proportion of known functions in LMI genomes, and thus in metagenomes where they thrive. While testing these hypotheses falls outside the scope of our work, we find the latter hypothesis more likely due to examples in existing literature that have successfully identified genes that belong to known metabolisms in some of the most obscure organisms via annotation strategies similar to those we have used in our work (Jaffe et al. 2020; Farag et al. 2020).

Taken together, these results (1) demonstrate that the PPCN normalization is an important consideration for investigations of metabolic enrichment in complex microbial communities as a function of microbial diversity, and (2) reveal that the enrichment of HMI populations in an environment offers a high resolution marker to resolve different levels of environmental stress.

Reference genomes with higher metabolic independence are over-represented in the gut metagenomes of individuals with IBD

So far, our findings demonstrate an overall, metagenome-level trend of increasing HMI within gut microbial communities as a function of IBD status without considering the individual genomes that contribute to this signal. Since the extent of metabolic independence of a microbial genome is a quantifiable trait, we considered a genome-based approach to validating our findings. Given the metagenome-level trends, we expected that the microbial genomes that encode a high number of metabolic modules associated with HMI should be more commonly detected in metagenomes from individuals diagnosed with IBD.

While publicly available reference genomes for microbial taxa will unlikely capture the diversity of individual gut metagenomes, we cast a broad net by surveying the ecology of 19,226 genomes in the Genome Taxonomy Database (GTDB) (Parks et al. 2022) that belonged to three major phyla associated with the human gut environment: Bacteroidetes, Firmicutes, and Proteobacteria (Woting and Blaut 2016; Turnbaugh et al. 2009). We used Human Microbiome Project data (Human Microbiome Project Consortium 2012) to characterize the distribution of these genomes across healthy human gut metagenomes. We used their single-copy core genes to identify genomes that were representative of microbial clades that are systematically detected in the healthy human gut (**Figure 3a**) and kept those that also occurred in at least 2% of samples from our set of 330 healthy and IBD metagenomes (see Methods). Selection of genomes that are relatively well-detected in the HMP dataset effectively removed taxa that primarily occur outside of the human gut. Of the final set of 338 reference genomes that passed our filters, 258 (76.3%) resolved to Firmicutes, 60 (17.8%) to Bacteroidetes, and 20 (5.9%) to Proteobacteria. Most of these genomes resolved to families common to the colonic microbiota, such as Lachnospiraceae (30.0%), Ruminococcaceae / Oscillospiraceae (23.1%), and Bacteroidaceae (10.1%) (Arumugam et al. 2011), while 5.9% belonged to poorly-studied families with temporary code names (Supplementary Table 3a). Finally, we performed a more comprehensive read recruitment analysis on this smaller set of genomes using all deeply-sequenced metagenomes from cohorts that included healthy, non-IBD, and IBD samples (**Figure 3**). This provided us with a quantitative summary of the detection patterns of GTDB genome representatives common to the human gut across our dataset.

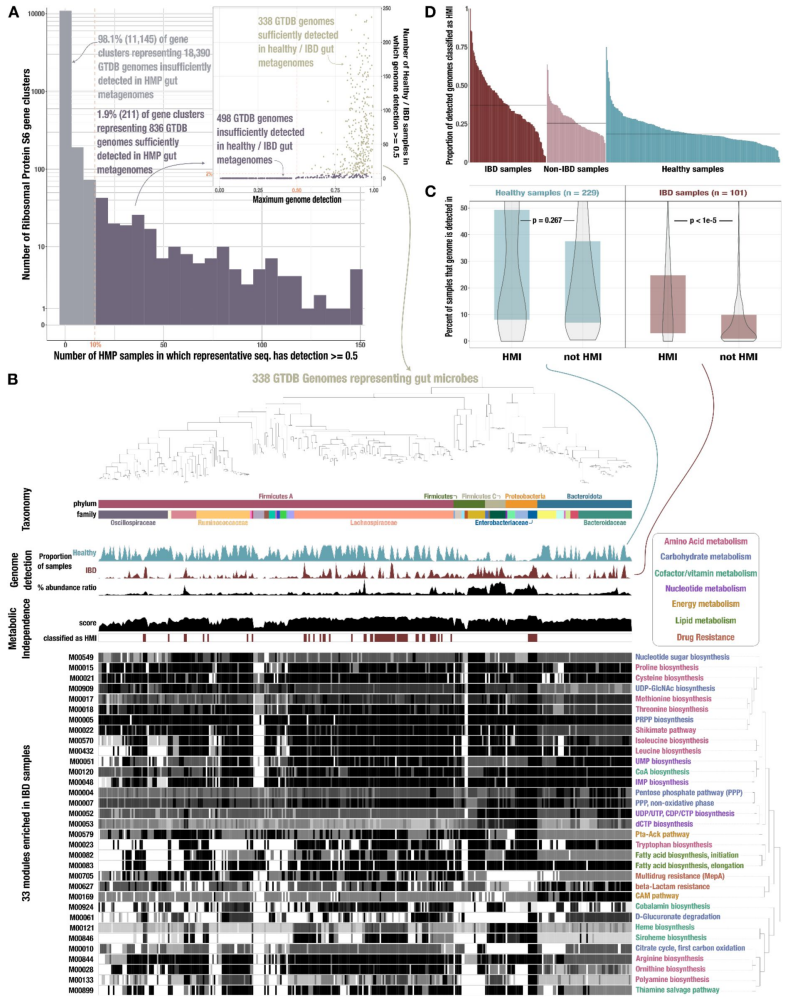


Figure 3.

Identification of HMI genomes and their distribution across gut samples.

A) Histogram of Ribosomal Protein S6 gene clusters (94% ANI) for which at least 50% of the representative gene sequence is covered by at least 1 read (>= 50% 'detection') in fecal metagenomes from the Human Microbiome Project (HMP) (Human Microbiome Project Consortium 2012). The dashed line indicates our threshold for reaching at least 50% detection in at least 10% of the HMP samples; gray bars indicate the 11,145 gene clusters that do not meet this threshold while purple bars indicate the 836 clusters that do. The subplot shows data for the 836 genomes whose Ribosomal Protein S6 sequences belonged to one of the passing (purple) gene clusters. The y-axis indicates the number of healthy/IBD gut metagenomes from our set of 330 in which the full genome sequence has at least 50% detection, and the x-axis indicates the genome's maximum detection across all 330 samples. The dashed line indicates our threshold for reaching at least 50% genome detection in at least 2% of samples; the 338 genomes that pass this threshold are tan and those that do not are purple. The phylogeny of these 338 genomes is shown in **B)** along with the following data, from top to bottom: taxonomic classification as assigned by GTDB; proportion of healthy samples with at least 50% detection of the genome sequence; proportion of IBD samples with at least 50% detection of the genome sequence; square-root normalized ratio of percent abundance in IBD samples to percent abundance in healthy samples; metabolic independence score (sum of completeness scores of 33 HMI-associated metabolic pathways); whether (red) or not (white) the genome is classified as having HMI with a threshold score of 26.4; heatmap of completeness scores for each of the 33 HMI-associated metabolic pathways (0% completeness is white and 100% completeness is black). Pathway name is shown on the right and colored according to its category of metabolism. **C)** Boxplot showing the proportion of healthy (blue) or IBD (red) samples in which genomes of each class are detected >= 50%, with p-values from a Wilcoxon Rank-Sum test on the underlying data. **D)** Barplot showing the proportion of detected genomes (with >= 50% genome sequence covered by at least 1 read) in each sample that are classified as HMI, for each group of samples. The black lines show the median for each group: 37.0% for IBD samples, 25.5% for non-IBD samples, and 18.4% for healthy samples.

We classified each genome as HMI if its average completeness of the 33 HMI-associated metabolic pathways was at least 80%, equivalent to a summed metabolic independence score of 26.4 (Methods). Across all genomes, the mean metabolic independence score was 24.0 (Q1: 19.9, Q3: 25.7). We identified 17.5% (59) of the reference genomes as HMI. HMI genomes were on average substantially larger (3.8 Mbp) than non-HMI genomes (2.9 Mbp) and encoded more genes (3,634 vs. 2,683 genes, respectively), which is in accordance with the reduced metabolic potential of non-HMI populations (Supplementary Table 3a). Our read recruitment analysis showed that HMI reference genomes were present in a significantly higher proportion of IBD samples compared to non-HMI genomes (**Figure 3c**, $p < 1e\text{-}5$, Wilcoxon Rank Sum test). Similarly, the fraction of HMI populations was significantly higher within a given IBD sample compared to samples classified as 'non-IBD' and those from healthy individuals (**Figure 3d**, $p < 1e\text{-}24$, Kruskal-Wallis Rank Sum test). In contrast, the detection of HMI populations and non-HMI populations was similar in healthy individuals (**Figure 3c**, $p = 0.267$, Wilcoxon Rank Sum test). The intestinal environment of healthy individuals likely supports both HMI and non-HMI populations, wherein 'metabolic diversity' is maintained by metabolic interactions such as cross feeding. Indeed, loss of cross-feeding interactions in the gut microbiome appears to be associated with a number of human diseases, including IBD (Marcelino et al. 2023). This interpretation is further supported by the fact that the top two HMI-associated pathways are required for the synthesis of cobalamin from glutamate. Auxotrophy for cobalamin biosynthesis is common among gut bacteria that rely on cross-feeding for this essential cofactor (Degnan et al. 2014; Magnúsdóttir et al. 2015; Kelly et al. 2019) (Supplementary File 1).

Overall, the classification of reference gut genomes as HMI and their enrichment in individuals diagnosed with IBD strongly supports the contribution of HMI to stress resilience of individual microbial populations. We note that survival in a disturbed gut environment will likely require a wide variety of additional functions that are not covered in the list of metabolic modules we consider to determine HMI status – for examples, see (Degnan et al. 2014; Martens et al. 2014; Zong et al. 2020; L. Feng et al. 2020; Goodman et al. 2009; Powell et al. 2016). Indeed, there may be many ways for a microbe to be metabolically independent, and our strategy likely failed to identify some HMI populations. Nonetheless, these data suggest that HMI serves as a reliable proxy for the identification of microbial populations that are particularly resilient.

HMI-associated metabolic potential predicts general stress on gut microbes

Our analysis identified HMI as an emergent property of gut microbial communities associated with individuals diagnosed with IBD. This community-level signal translates to individual microbial populations and provides insights into the microbial ecology of stressed gut environments. HMI-associated metabolic pathways were enriched at the community level, and microbial populations encoding these modules were more prevalent in individuals with IBD than in healthy individuals. Furthermore, the copy number of these pathways and the proportion of HMI populations reflect the severity of environmental stress and translate to host health states (**Supplementary Figure 5b**, **Figure 3d**). The ecological implications of these observations suggest that HMI may serve as a predictor of general stress in the human gut environment.

So far, efforts to identify IBD using microbial markers have presented classifiers based on (1) taxonomy in pediatric IBD patients (Papa et al. 2012; Gevers et al. 2014), (2) community composition in combination with clinical data (Halfvarson et al. 2017), (3) untargeted metabolomics and/or species-level relative abundance from metagenomes (Franzosa et al. 2019) and (4) k-mer-based sequence variants in metagenomes that can be linked to microbial genomes associated with IBD (Reiter et al. 2022). Performance varied both between and within studies according to the target classes and data types used for training and validation of each classifier (Supplementary Table 4a). For those studies reporting accuracy, a maximum accuracy of 77% was achieved based on either metabolite profiles (for prediction of IBD-subtype) (Franzosa et al.

2019) or k-mer-based sequence variants (for differentiating between IBD and non-IBD samples) (Reiter et al. 2022). Some studies reported performance as area under the receiver operating characteristic curve (AUROCC), a typical measure of classifier utility describing both sensitivity (ability to correctly identify the disease) and specificity (ability to correctly identify absence of disease). For this metric the highest value was 0.92, achieved by (Franzosa et al. 2019) when using metabolite profiles, with or without species abundance data, for classifying IBD vs non-IBD. However, the majority of these classifiers were trained and tested on a relatively small group of individuals that all come from the same region, i.e. clinical studies confined to a specific hospital. Though some had high performance, they either relied on data that are inaccessible to most laboratories and clinics considering that untargeted metabolomics analyses are difficult to reproduce (Koek et al. 2011; Lin et al. 2020), or they required complex k-mer-based models without the resolution to differentiate gradients in host health (Reiter et al. 2022). These classifiers thus have limited translational potential across global clinical settings and do not provide an ecological framework to explain the observed shifts in community composition and activity. For practical use as a diagnostic tool, a microbiome-based classifier for IBD should rely on an ecologically meaningful, easy to measure, and high-level signal that is robust to host variables like lifestyle, geographical location, and ethnicity. High metabolic independence could potentially fill this gap as a metric related to the ecological filtering that defines microbial community changes in the IBD gut microbiome.

We trained a logistic regression classifier to explore the applicability of HMI as a non-invasive diagnostic tool for IBD. The classifier's predictors were the per-population copy numbers of IBD-enriched metabolic pathways in a given metagenome. Across the 330 deeply-sequenced IBD and healthy samples included in this analysis, the classifier had high sensitivity and specificity (**Figure 4**). It correctly identified (on average) 76.8% of samples from individuals diagnosed with IBD and 89.5% of samples representing healthy individuals, for an overall accuracy of 85.6% and an average AUROCC of 0.832 (Supplementary Table 4c). Our model outperforms (Gevers et al. 2014; Halfvarson et al. 2017; Reiter et al. 2022) or has comparable performance to (Franzosa et al. 2019; Papa et al. 2012) the previous attempts to classify IBD from fecal samples in more restrictively-defined cohorts. It also has the advantage of being a simple model, utilizing a relatively low number of features compared to the other classifiers. Thus, HMI shows promise as an accessible diagnostic marker of IBD. Due to the lack of time-series studies that include individuals in the pre-diagnosis phase of IBD development, we cannot test the applicability of HMI to predict IBD onset (Lloyd-Price et al. 2019).

Yet, the gradient of metabolic independence reflected by per-population pathway copy number and the relative increase in the number of HMI populations detected in non-IBD samples (**Supplementary Figure 5b**, **Figure 3d**) suggests that the degree of HMI in the gut microbiome may be indicative of general gut stress, such as the stress induced by antibiotic use. Antibiotics can cause long-lasting perturbations of the gut microbiome – including reduced diversity, emergence of opportunistic pathogens, increased microbial load, and development of highly-resistant strains – with potential implications for host health (Ramirez et al. 2020). We applied our metabolism classifier to a metagenomic dataset that reflects the changes in the microbiome of healthy people before, during and up to 6 months following a 4-day antibiotic treatment (Palleja et al. 2018). The resulting pattern of sample classification corresponds to the post-treatment decline and subsequent recovery of species richness documented in the study by (Palleja et al. 2018).

All pre-treatment samples were classified as 'healthy' followed by a decline in the proportion of 'healthy' samples to a minimum 8 days post-treatment, and a gradual increase until 180 days post treatment, when over 90% of samples were classified as 'healthy' (**Figure 5**, Supplementary Table 4b). These observations support the role of HMI as an ecological driver of microbial resilience during gut stress caused by a variety of environmental perturbations and demonstrate its diagnostic power in reflecting gut microbiome state.

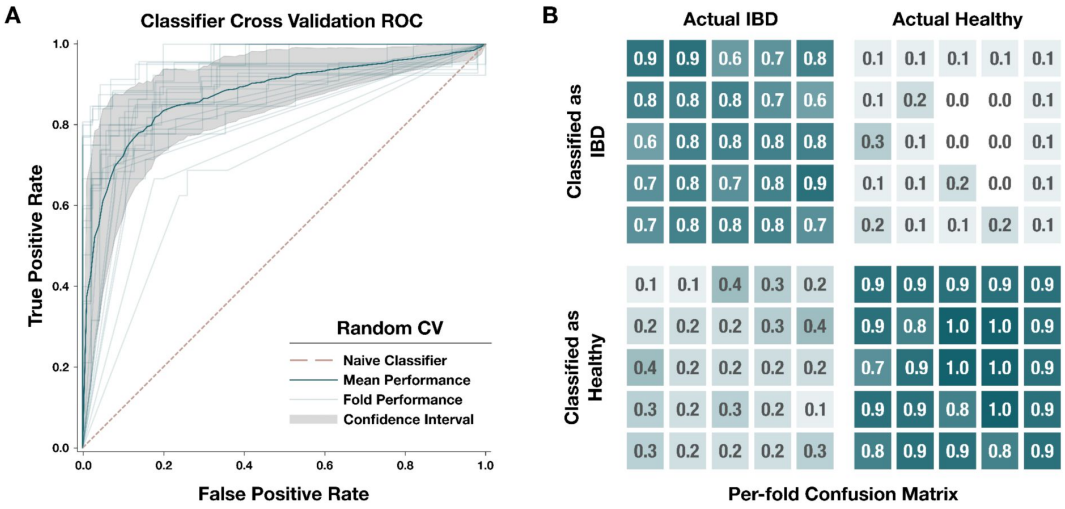


Figure 4.

Performance of our metagenome classifier trained on per-population copy numbers of IBD-enriched modules.

A) Receiver operating characteristic (ROC) curves for 25-fold cross-validation. Each fold used a random subset of 80% of the data for training and the other 20% for testing. In each fold, we calculated a set of IBD-enriched modules from the training dataset and used the PPCN of these modules to train a logistic regression model whose performance was evaluated using the test dataset. Light gray lines show the ROC curve for each fold, the dark blue line shows the mean ROC curve, the gray area delineates the confidence interval for the mean ROC, and the pink dashed line indicates the benchmark performance of a naive (random guess) classifier. **B)** Confusion matrix for each fold of the random cross-validation. Categories of classification, from top left to bottom right, are: true positives (correctly classified IBD samples), false positives (incorrectly classified Healthy samples), false negatives (incorrectly classified IBD samples), and true negatives (correctly classified Healthy samples). Each fold is represented by a box within each category. Opacity of the box indicates the proportion of samples in that category, and the actual proportion is written within the box with one significant digit. Underlying data for this matrix can be accessed in Supplementary Table 4d.

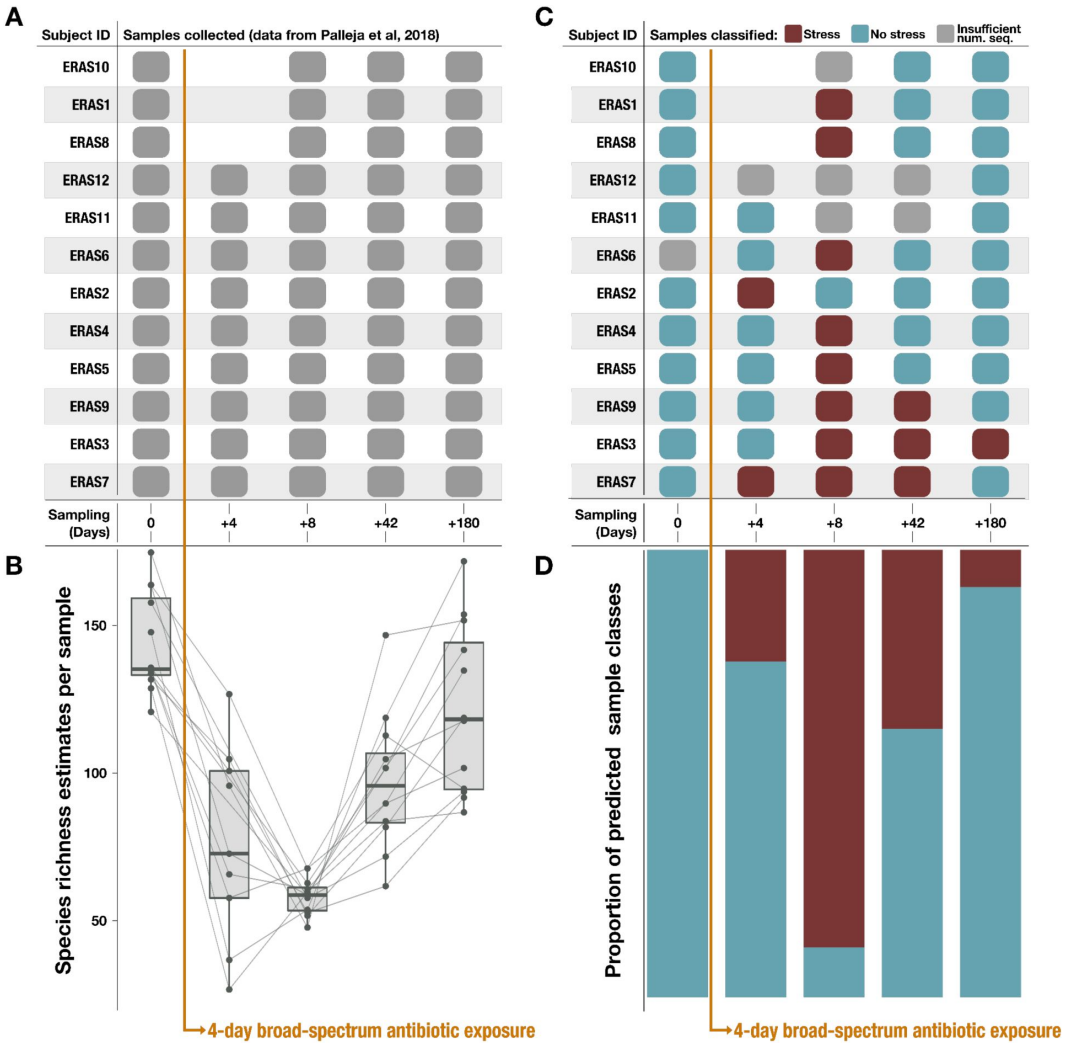


Figure 5.

Classification results on an antibiotic time-series dataset from

(Palleja et al. 2018). Note that antibiotic treatment was taken on days 1-4. **A)** Samples collected per subject during the time series. **B)** Species richness data (figure reproduced from (Palleja et al. 2018)). **C)** Classification of each sample by the metabolism classifier profiled in **Figure 4**. Samples with insufficient sequencing depth were not classified. **D)** Proportion of classes assigned to samples per day in the time series.

Conclusions

Overall, our observations that stem from the analysis of hundreds of reference genomes, deeply-sequenced gut metagenomes, and multiple categories of human disease states suggest that environmental stress in the human gut – whether it is associated with inflammation, cancer, or antibiotic use – promotes the survival and relative expansion of microbial populations with high metabolic independence. These results establish HMI as a high-level metric to classify gradients of human health states through the gut microbiota that is robust to ethnic, geographical or lifestyle factors. Taken together with recent evidence that models altered ecological relationships within gut microbiomes under stress due to disrupted metabolic cross-feeding (Heinken, Hertel, and Thiele 2021; Marcelino et al. 2023), our data support the hypothesis that the reduction in microbial diversity, or more generally 'dysbiosis', is an emergent property of microbial communities responding to disease pathogenesis or other external factors such as antibiotic use that disrupt the gut microbial ecosystem. This paradigm depicts microbes as bystanders by default, rather than perpetrators or drivers of noncommunicable human diseases, and provides an ecological framework to explain the frequently observed reduction in microbial diversity associated with IBD and other noncommunicable human diseases and disorders.

Methods

A bioinformatics workflow that further details all analyses described below and gives access to reproducible data products is available at the URL *https://merenlab.org/data/ibd-gut-metabolism/*.

A new framework for metabolism estimation

We developed a new program 'anvi-estimate-metabolism' (*https://anvio.org/m/anvi-estimate-metabolism*), which uses gene annotations to estimate 'completeness' and 'copy number' of metabolic pathways that are defined in terms of enzyme accession numbers. By default, this tool works on metabolic modules from the KEGG MODULE database (Kanehisa et al. 2012, 2023) which are defined by KEGG KOfams (Aramaki et al. 2019), but user-defined modules based on a variety of functional annotation sources are also accepted as input. Completeness estimates describe the percentage of steps (typically, enzymatic reactions) in a given metabolic pathway that are encoded in a genome or a metagenome. Likewise, copy number summarizes the number of distinct sets of enzyme annotations that collectively encode the complete pathway. This program offers two strategies for estimating metabolic potential: a 'stepwise' strategy with equivalent treatment for alternative enzymes – i.e, enzymes that can catalyze the same reaction in a given metabolic pathway – and a 'pathwise' strategy that accounts for all possible variations of the pathway. The Supplementary File 1 file includes more information on these two strategies and the completeness/copy number calculations. For the analysis of metagenomes, we used stepwise copy number of KEGG modules. Briefly, the calculation of stepwise copy number is done as follows: the copy number of each step in a pathway (typically, one chemical reaction or conversion) is individually evaluated by translating the step definition into an arithmetic expression that summarizes the number of annotations for each required enzyme. In cases where multiple enzymes or an enzyme complex are needed to catalyze the reaction, we take the minimum number of annotations across these components. In cases where there are alternative enzymes that can each catalyze the reaction individually, we sum the number of annotations for each alternative. Once the copy number of each step is computed, we then calculate the copy number of the entire pathway by taking the minimum copy number across all the individual steps. The use of minimums results in a conservative estimate of pathway copy number such that only copies of the pathway with all enzymes present are counted. For the analysis of genomes, we calculated the

stepwise completeness of KEGG modules. This calculation is similar to the one described above for copy number, except that the step definition is translated into a boolean expression that, once evaluated, indicates the presence or absence of each step in the pathway. Then, the completeness of the modules is computed as the proportion of present steps in the pathway.

Metagenomic Datasets and Sample Groups

We acquired publicly-available gut metagenomes from 13 different studies (Le Chatelier et al. 2013; Q. Feng et al. 2015; Franzosa et al. 2019; Lloyd-Price et al. 2019; Qin et al. 2012; Quince et al. 2015; Rampelli et al. 2015; Raymond et al. 2016; Schirmer et al. 2018; Vineis et al. 2016; "BioProject," n.d.; Wen et al. 2017; Xie et al. 2016). The studies were chosen based on the following criteria: (1) they included shotgun metagenomes of fecal matter (primarily stool, but some ileal pouch luminal aspirate samples (Vineis et al. 2016) are also included); (2) they sampled from people living in industrialized countries (in the case where a study (Rampelli et al. 2015) included samples from hunter-gatherer populations, only the samples from industrialized areas were included in our analysis); (3) they included samples from people with IBD and/or they included samples from people without gastrointestinal (GI) disease or inflammation; and (4) clear metadata differentiating between case and control samples was available. A full description of the studies and samples can be found in Supplementary Table 1a-c. We grouped samples according to the health status of the sample donor. Briefly, the 'IBD' group of samples includes those from people diagnosed with Crohn's disease (CD), ulcerative colitis (UC), or pouchitis. The 'non-IBD' group contains non-IBD controls, which includes both healthy people presenting for routine cancer screenings as well as people with benign or non-specific symptoms that are not clinically diagnosed with IBD. Colorectal cancer patients from (Q. Feng et al. 2015) were also put into the 'non-IBD' group on the basis that tumors in the GI tract may arise from local inflammation (Kraus and Arber 2009) and represent a source of gut stress without an accompanying diagnosis of IBD. Finally, the 'HEALTHY' group contains samples from people without GI-related diseases or inflammation. Note that only control or pre-treatment samples were taken from the studies covering type 2 diabetes (Qin et al. 2012), ankylosing spondylitis (Wen et al. 2017), antibiotic treatment (Raymond et al. 2016), and dietary intervention ("BioProject," n.d.); these controls were all assigned to the 'HEALTHY' group. At least one study (Le Chatelier et al. 2013) included samples from obese people, and these were also included in the 'HEALTHY' group.

Processing of metagenomes

We made single assemblies of most gut metagenomes using the anvi'o metagenomics workflow implemented in the program 'anvi-run-workflow' (Shaiber et al. 2020). This workflow uses Snakemake (Köster and Rahmann 2012), and a tutorial is available at the URL *https://merenlab.org/2018/07/09/anvio-snakemake-workflows/*. Briefly, the workflow includes quality filtering using 'iu-filter-quality-minochè (Eren et al. 2013); assembly with IDBA-UD (Peng et al. 2012) (using a minimum contig length of 1000); gene calling with Prodigal v2.6.3 (Hyatt et al. 2010); tRNA identification with tRNAscan-SE v2.0.7 (Chan and Lowe 2019); and gene annotation of ribosomal proteins (Seemann, n.d.), single-copy core gene sets (M. D. Lee 2019), KEGG KOfams (Aramaki et al. 2019), NCBI COGs (Galperin et al. 2021), and Pfam (release 33.1, (Mistry et al. 2021)). The aforementioned annotation was done with programs that relied on HMMER v3.3.2 (Eddy 2011) as well as Diamond v0.9.14.115 (Buchfink, Xie, and Huson 2015). As part of this workflow, all single assemblies were converted into anvi'o contigs databases. Samples from (Vineis et al. 2016) were processed differently because they contained merged reads rather than individual paired-end reads: no further quality filtering was run on these samples, we assembled them individually using MEGAHIT (Li et al. 2015), and we used the anvi'o contigs workflow to perform all subsequent steps described for the metagenomics workflow above. Note that we used a version of KEGG downloaded in December 2020 (for reproducibility, the hash of the KEGG snapshot available via 'anvi-setup-kegg-kofams' is 45b7cc2e4fdc). Additionally, the annotation

program ‘anvi-run-kegg-kofams‘ includes a heuristic for annotating hits with bitscores that are just below the KEGG-defined threshold, which is described at *https://anvio.org/m/anvi-run-kegg-kofams/*.

Genomic Dataset

We also analyzed microbial genomes from the Genome Taxonomy Database (GTDB), release 95.0 (Parks et al. 2018, 2020). We downloaded all reference genome sequences for the species cluster representatives.

Processing of GTDB genomes

We converted all GTDB genomes into anvi’o contigs databases and annotated them using the anvi’o contigs workflow, which is similar to the metagenomics workflow described above and uses the same programs for gene identification and annotation.

Estimation of the number of microbial populations per metagenome

We used single-copy core gene (SCG) sets belonging to each domain of microbial life (Bacteria, Archaea, Protista) (M. D. Lee 2019) to estimate the number of populations from each domain present in a given metagenomic sample. For each domain, we calculated the number of populations by taking the mode of the number of copies of each SCG in the set. We then summed the number of populations from each domain to get a total number of microbial populations within each sample. We accomplished this using SCG annotations provided by ‘anvi-run-hmms‘ (which was run during metagenome processing) and a custom script relying on the anvi’o class ‘NumGenomesEstimator‘ (see reproducible workflow).

Removal of samples with low sequencing depth

We observed that, at lower sequencing depths, our estimates for the number of populations in a metagenomic sample were moderately correlated with sequencing depth (**Supplementary Figure 1**, $R > 0.5$). These estimates rely on having accurate counts of single-copy core genes (SCGs), so we hypothesized that lower-depth samples were systematically missing SCGs, especially from populations with lower abundance. Since accurate population number estimates are critical for proper normalization of pathway copy numbers, keeping these lower-depth samples would have introduced a bias into our metabolism analyses. To address this, we removed samples with low sequencing depth from downstream analyses using a sequencing depth threshold of 25 million reads, such that the remaining samples exhibited a weaker correlation ($R < 0.5$) between sequencing depth and number of estimated populations. We kept samples for which both the R1 file and the R2 file contained at least 25 million reads (and for the (Vineis et al. 2016) dataset, we kept samples containing at least 25 million merged reads). This produced our final sample set of 408 metagenomes.

Estimation of normalized pathway copy numbers in metagenomes

We ran ‘anvi-estimate-metabolism‘, in genome mode and with the ‘--add-copy-number‘ flag, on each individual metagenome assembly to compute stepwise copy numbers for KEGG modules from the combined gene annotations of all populations present in the sample. We then divided these copy numbers by the number of estimated populations within each sample to obtain a per-population copy number (PPCN) for each pathway.

Selection of IBD-enriched pathways

We used a one-sided Mann-Whitney-Wilcoxon test with a FDR-adjusted p-value threshold of p <= 2e-10 on the per-sample PPCN values for each module individually to identify the pathways that were most significantly enriched in the IBD sample group compared to the healthy group. We calculated the median per- population copy number of each metabolic pathway in the IBD samples, and again in the healthy samples. After filtering for p-values <= 2e-10, we also applied a minimum effect size threshold based on the median per-population copy number in each group (M_{IBD} - $M_{Healthy}$ >= 0.12) – this threshold was calculated by taking the mean effect size over all pathways that passed the p-value threshold. The set originally contained 34 pathways that passed both thresholds, but we removed one redundant module (M00006) which represents the first half of another module in the set (M00004).

Test for enrichment of biosynthesis pathways

We used a one-sided Fisher's exact test (also known as hypergeometric test, see e.g., (Boyle et al. 2004)) for testing the independence between the metabolic pathways identified to be IBD-enriched (i.e., using the methods described in "Selection of IBD-enriched pathways) and functionality (i.e., pathways annotated to be involved in biosynthesis).

Pathway comparisons

Because the 33 IBD-enriched pathways were selected using PPCNs of healthy and IBD samples, statistical tests comparing PPCN distributions for these modules need to be interpreted with care, because the hypotheses were selected and tested on the same dataset (Fithian, Sun, and Taylor 2014). Therefore, to assess the statistical validity of the identified IBD-enriched modules, we performed the following repeated sample-split analysis: we first randomly split the IBD and healthy samples into the equal-sized training and validation sets. We select IBD-enriched modules in the training set using the Mann-Whitney-Wilcoxon test, and then compute the p-values on the validation set. We repeat this sample split analysis 1,000 times with an FDR-adjusted p-value threshold of 1e-10 on the first split; most identified modules (89.4%; 95% CI: [87.5%, 91.3%]) on the training sets remain significant at a slightly less stringent threshold (1e-8) on the validation sets. This indicates that the approach we used to identify IBD-enriched modules yields stable and statistically significant results on this dataset.

Metagenome classification

We trained logistic regression models to classify samples as 'IBD' or 'healthy' using per-population copy numbers of IBD-enriched modules as features. We ran a 25-fold cross-validation pipeline on the set of 330 healthy and IBD metagenomes in our analysis, using an 80% train – 20% test random split of the data in each fold. The pipeline included selection of IBD-enriched pathways within the training samples using the same strategy as described above, followed by training and testing of a logistic regression model as implemented in the 'sklearn' Python package. We set the 'penalty' parameter of the model to "None" and the 'max_iter' parameter to 20,000 iterations, and we used the same random state in each fold to ensure changes in performance only come from differences in the training data rather than differences in model initialization. To summarize the overall performance of the classifier, we took the mean (over all folds) of each performance metric.

We trained a final classifier using the 33 IBD-enriched pathways selected earlier from the entire set of 330 healthy and IBD metagenomes. We then applied this classifier to the metagenomic samples from (Palleja et al. 2018), which we processed in the same way as the other samples in our analysis (including removal of samples with low sequencing depth and calculation of PPCNs of KEGG modules for use as input features to the classifier model).

Identification of gut microbial genomes from the GTDB

We took 19,226 representative genomes from the GTDB species clusters belonging to the phyla Firmicutes, Bacteroidetes, and Proteobacteria, which are most common in the human gut microbiome (Woting and Blaut 2016). To evaluate which of these genomes might represent gut microbes in a computationally-tractable manner, we ran the anvi'o 'EcoPhylo' workflow (*https://anvio.org/m/ecophylo*) to contextualize these populations within 150 healthy gut metagenomes from the Human Microbiome Project (HMP) (Human Microbiome Project Consortium 2012). Briefly, the EcoPhylo workflow (1) recovers sequences of a gene family of interest from each genome and metagenomic sample in the analysis, (2) clusters resulting sequences and picks representative sequences using mmseqs2 (Steinegger and Söding 2017), and (3) uses the representative sequences to rapidly summarize the distribution of each population cluster across the metagenomic samples through metagenomic read recruitment analyses. Here, we used the Ribosomal Protein S6 as our gene of interest, since it was the most frequently-assembled single-copy-core gene in our set of GTDB genomes. We clustered the Ribosomal Protein S6 sequences from GTDB genomes at 94% nucleotide identity.

To identify genomes that were likely to represent gut microbes, we selected genomes whose ribosomal protein S6 belonged to a gene cluster where at least 50% of the representative sequence was covered (i.e. detection >= 0.5x) in more than 10% of samples (i.e. n > 15). There are 100 distinct individuals represented in the 150 HMP gut metagenomes – 56 of which were sampled just once and 46 of which were sampled at 2 or 3 time points – so this threshold is equivalent to detecting the genome in 5% - 15% of individuals. From this selection we obtained a set of 836 genomes; however, these were not exclusively gut microbes, as some non-gut populations have similar ribosomal protein S6 sequences to gut microbes and can therefore pass this selection step. To eliminate these, we mapped our set of 330 healthy and IBD metagenomes to the 836 genomes using the anvi'o metagenomics workflow, and extracted genomes whose entire sequence was at least 50% covered (i.e. detection >= 0.5x) in over 2% (n > 6) of these samples. Our final set of 338 genomes was used in downstream analysis.

Genome phylogeny

To create the phylogeny, we identified the following ribosomal proteins that were annotated in at least 90% (n = 304) of the genomes: Ribosomal_S6, Ribosomal_S16, Ribosomal_L19, Ribosomal_L27, Ribosomal_S15, Ribosomal_S20p, Ribosomal_L13, Ribosomal_L21p, Ribosomal_L20, and Ribosomal_L9_C. We used 'anvi-get-sequences-for-hmm-hits' to extract the amino acid sequences for these genes, align the sequences using MUSCLE v3.8.1551 (Edgar 2004), and concatenate the alignments. We used trimAl v1.4.rev15 (Capella-Gutiérrez, Silla-Martínez, and Gabaldón 2009) to remove any positions containing more than 50% of gap characters from the final alignment. Finally, we built the tree with IQtree v2.2.0.3 (Minh et al. 2020), using the WAG model and running 1,000 bootstraps.

Determination of HMI status for genomes

We estimated metabolic potential for each genome with 'anvi-estimate-metabolism' (in genome mode) to get stepwise completeness scores for each KEGG module, and then we used the script 'anvi-script-estimate-metabolic-independencè to give each genome a metabolic independence score based on completeness of the 33 IBD-enriched pathways. Briefly, the latter script calculates the score by summing the completeness scores of each pathway of interest. Genomes were classified as having high metabolic independence (HMI) if their score was greater than or equal to 26.4. We calculated this threshold by requiring these 33 pathways to be, on average, at least 80% complete in a given genome.

Genome distribution across sample groups

We mapped the gut metagenomes from the healthy, non-IBD, and IBD groups to each genome using the anvi'o metagenomics workflow in reference mode. We used 'anvi-summarizè to obtain a matrix of genome detection across all samples. We summarized this data as follows: for each genome, we computed the proportion of samples in each group in which at least 50% of the genome sequence was covered by at least 1 read (>= 50% detection). For each sample, we calculated the proportion of detected genomes that were classified as HMI. We also computed the percent abundance of each genome in each sample by dividing the number of reads mapping to that genome by the total number of reads in the sample.

Visualizations

We used ggplot2 (Wickham 2016) to generate most of the initial data visualizations. The phylogeny and heatmap in **Figure 3** were generated by the anvi'o interactive interface and the ROC curves in **Figure 4** were generated using the pyplot package of matplotlib (Hunter 2007). These visualizations were refined for publication using Inkscape, an open-source graphical editing software that is available at *https://inkscape.org/*.

Data Availability

Accession numbers for publicly available data are listed in our Supplementary Tables at doi:*10.6084/m9.figshare.22679080*. Our Supplementary Files are also available at doi:*10.6084/m9.figshare.22679080*. Contigs databases of our assemblies for the 408 deeply-sequenced metagenomes can be accessed at doi:*10.5281/zenodo.7872967*, and databases for our assemblies of the (Palleja et al. 2018) metagenomes can be accessed at doi:*10.5281/zenodo.7897987*. Contigs databases of the 338 GTDB gut reference genomes are available at doi:*10.5281/zenodo.7883421*.

Acknowledgements

We thank Chris Quince for advice on statistical significance testing. IV acknowledges support from the National Science Foundation Graduate Research Fellowship under Grant No. 1746045; ADW acknowledges support from the National Institutes of General Medical Sciences under R35 GM133420. YTC acknowledges support from the Stanford Data Science Postdoctoral Fellowship. RB acknowledges support from the National Institutes of Health under R35 GM128716. ECF acknowledges support from the University of Chicago International Student Fellowship. Additional support for ECF and AME came from an NIH NIDDK grant (RC2 DK122394) to AME.

Additional information

Author Contributions

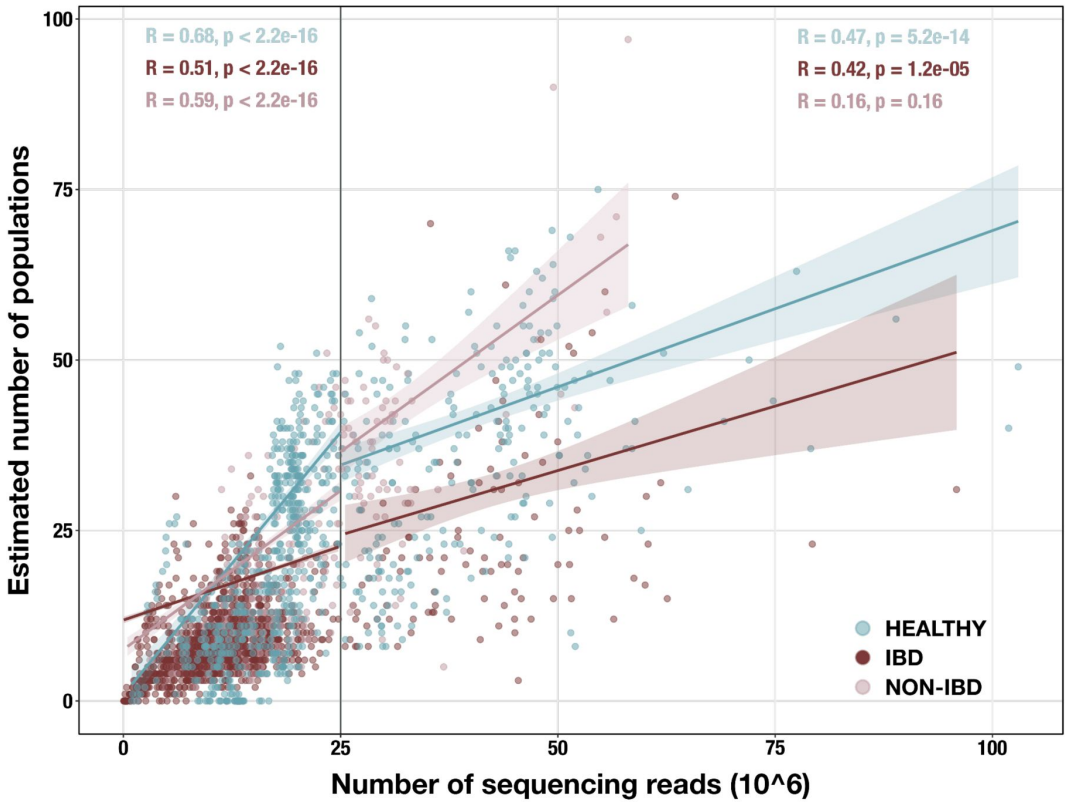
IV, JF and AME conceived the study. IV developed methodology. IV, MSS and AME developed computational analysis tools. IV, JF, and AME performed formal analyses. YTC supported statistical analyses. RB, BJ, ADW and MKY conducted investigations. ECF, ARW, CV and AFG provided resources. ECF, ARW and IV curated data. IV and AME prepared figures. IV, JF and AME wrote the paper with critical input from all authors. JF and AME supervised the project.

Ethics declarations

Competing interests

The authors declare that they have no competing interests.

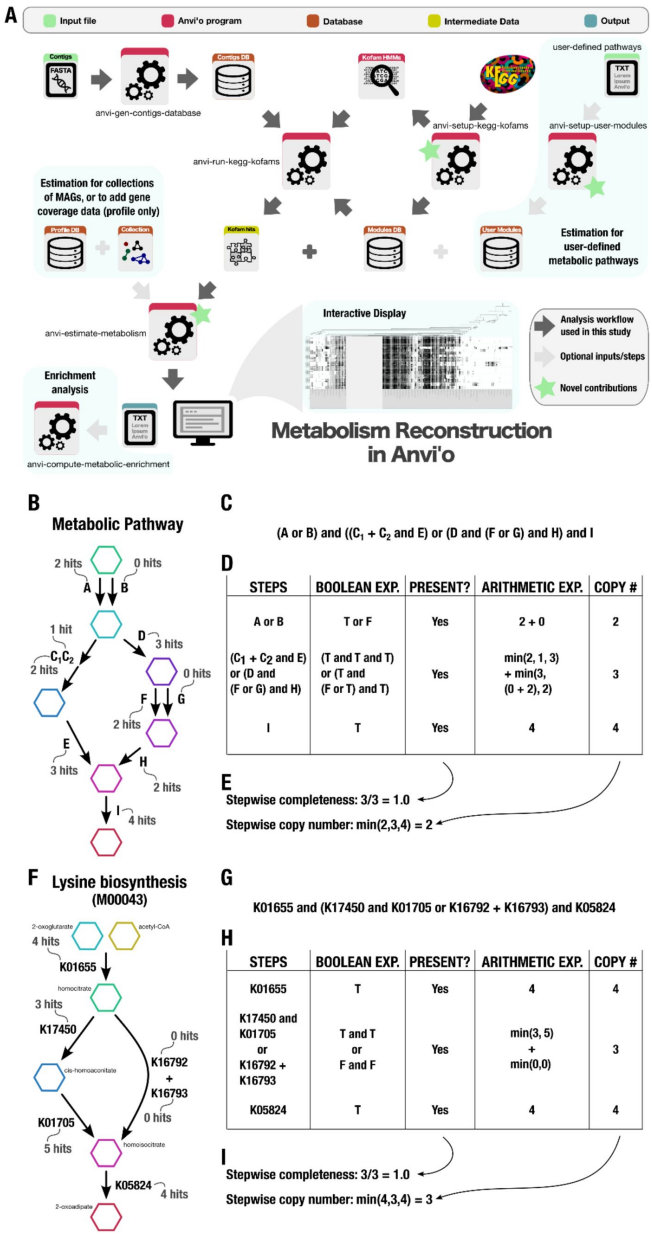
Supplemental Figures



Supplementary Figure 1.

Scatterplot of sequencing depth vs estimated number of microbial populations in each of 2,893 stool metagenomes.

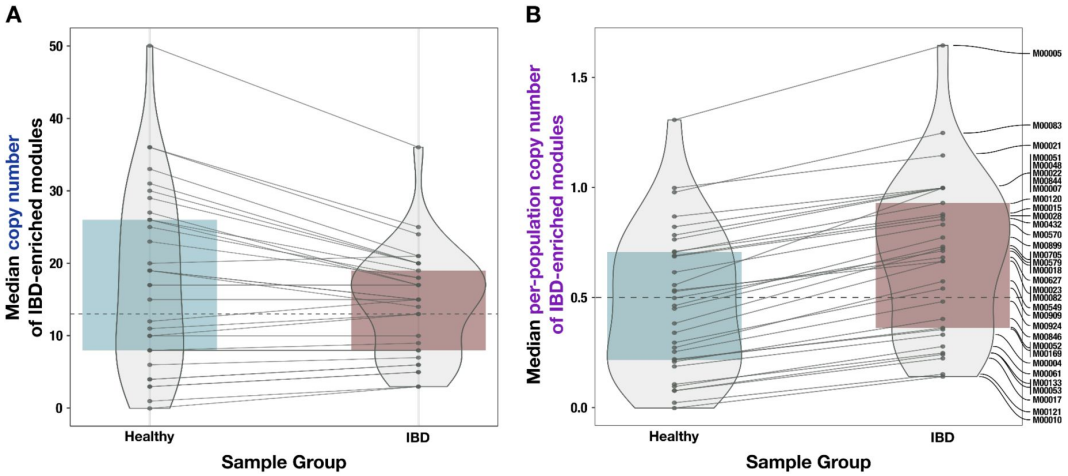
Sequencing depth is represented by the number of R1 reads, except for (Vineis et al. 2016) samples, in which case it is the number of merged paired-end reads. The vertical line indicates our sequencing depth threshold of 25 million reads. Per-group Spearman's correlation coefficients and p-values are shown for the subset of samples with depth < 25 million reads (top left) and for the subset with depth >= 25 million reads (top right). Regression lines are shown for each group in each subset, with standard error indicated by the colored background.



Supplementary Figure 2.

Technical details of the metabolism reconstruction software framework in anvi'o. A)

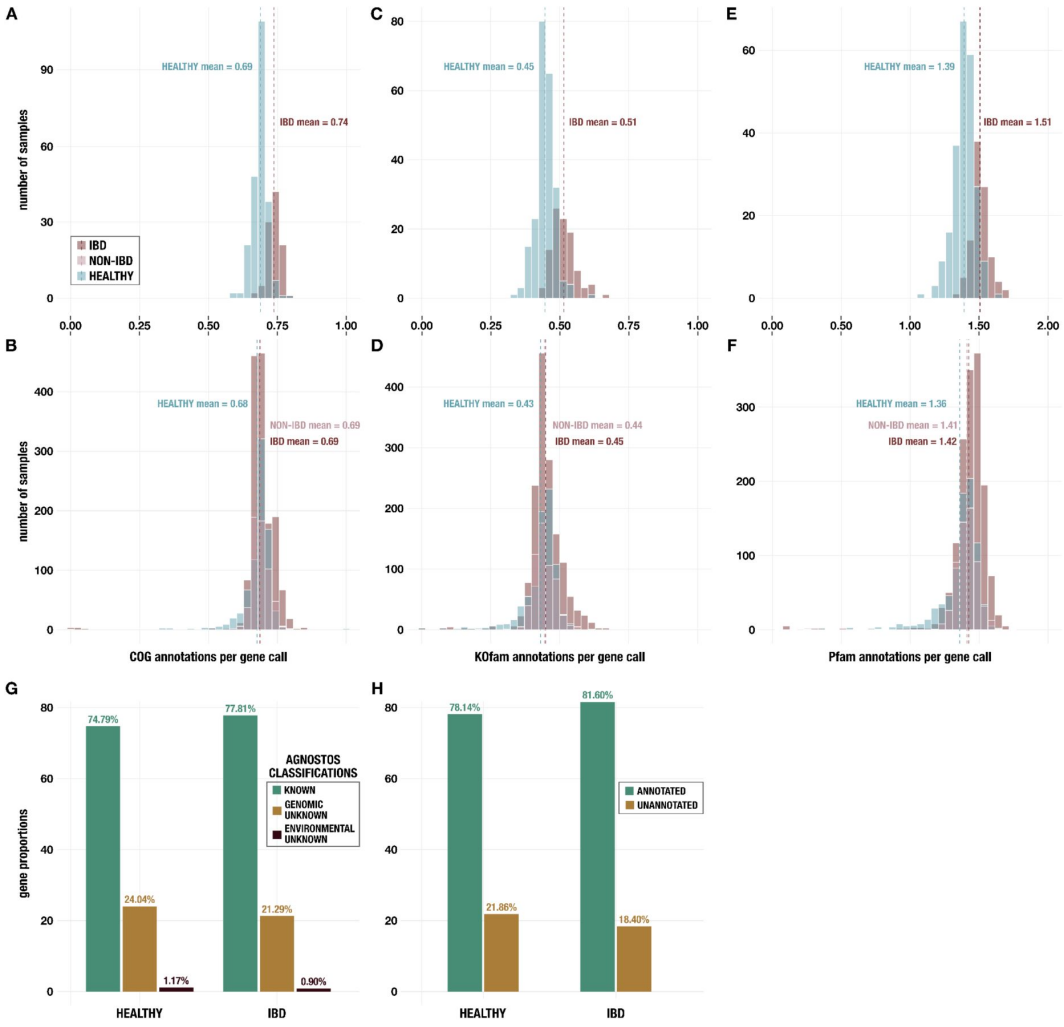
Workflow of metabolism reconstruction programs and their inputs/outputs. Dark arrows indicate the primary analysis path utilized in this study. Blue background indicates optional features in the framework. A demonstration of completeness score and copy number calculations for metabolic pathways (performed by the program 'anvi-estimate-metabolism' is shown using example enzyme annotation data in panels B – E (for a theoretical pathway) and F – I (for a real pathway). **B)** Theoretical metabolic pathway, where hexagons represent metabolites, arrows represent chemical reactions, letters represent enzymes (subscripts indicate enzyme components), and the example number of gene annotation hits for each enzyme is written in gray. **C)** The definition of the theoretical pathway from panel B, written in terms of the required enzymes. **D)** Table showing the major steps in the pathway and example calculations for step presence and copy number. Step presence is calculated by evaluating a boolean expression created from the step definition in which enzymes with > 0 hits are replaced with True (T) and the others with False (F). Step copy number is calculated by evaluating the corresponding arithmetic expression in which the enzymes are replaced with their annotation counts. **E)** Final calculations of completeness score (fraction of present steps) and copy number for the theoretical metabolic pathway. **F – I)** Same as panels B – E, but for KEGG module M00043.



Supplementary Figure 3.

Comparison of unnormalized copy number data and normalized (per-population copy number, or PPCN) data for the IBD-enriched modules.

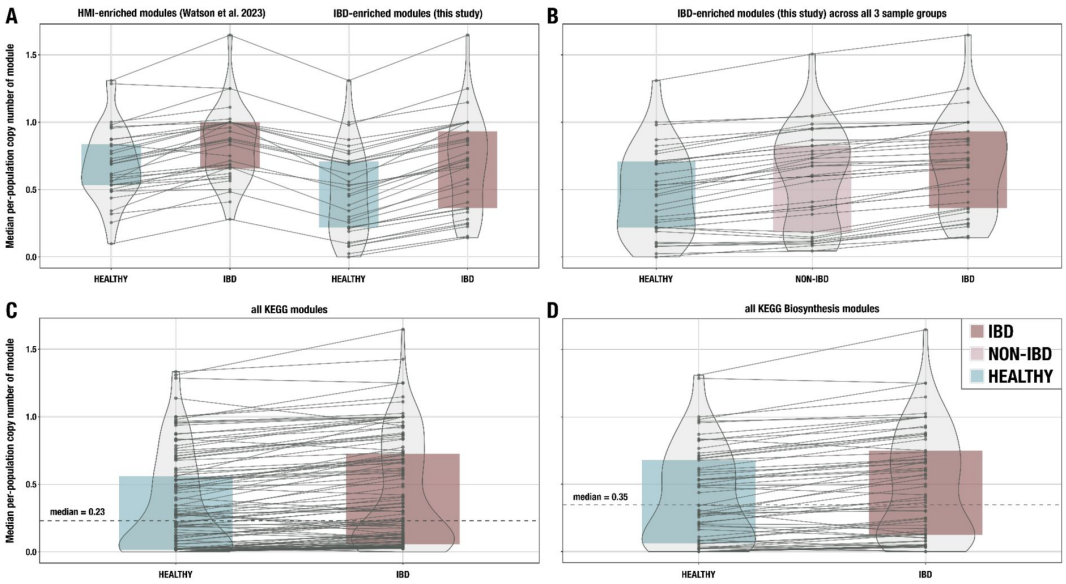
A) Boxplot of median copy numbers for each module in the healthy samples (blue) and IBD samples (red). **B)** Boxplots of median PPCN for each module in the healthy samples (blue) and IBD samples (red). Lines connect data points for the same module in each plot. The gray dashed line in each plot indicates the overall median value.



Supplementary Figure 4.

Histograms of annotations per gene call

from **A,B)** NCBI COGs; **C,D)** KEGG KOfams; and **E,F)** Pfams. Panels A, C, and E show data for metagenomes in the subset of 330 deeply-sequenced samples from healthy people and people with IBD, and panels B, D, and F show data for all 2,893 samples including those from non-IBD controls. **G)** Proportion of genes with each classification from AGNOSTOS (Vanni et al. 2022) in the subset of 330 deeply-sequenced samples. **H)** Proportion of genes with at least one annotation from KEGG KOfams (Aramaki et al. 2020), NCBI COGs (Galperin et al. 2015), or Pfams (Mistry et al. 2021) (green) and proportion without any annotation (brown) in the subset of 330 deeply-sequenced samples.



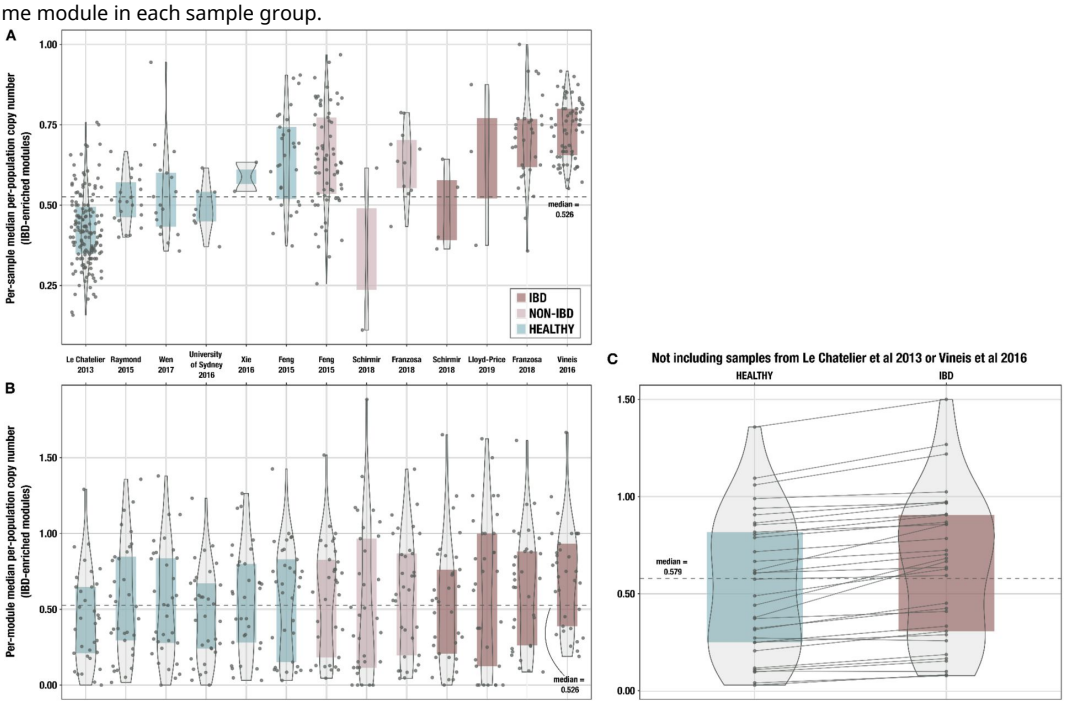
Supplementary Figure 5.

Additional boxplots of median per-population copy number for various subsets of metabolic pathways and metagenome samples. A)

33 modules enriched in HMI populations from (Watson et al. 2023) compared to the 33 IBD-enriched modules from this study, with medians computed in the set of deeply-sequenced healthy (n = 229) and IBD (n = 101) samples. **B)** The 33 IBD-enriched modules from this study, with medians computed in the set of deeply-sequenced healthy (n = 229), non-IBD (n = 78), and IBD (n = 101) samples. **C)** All KEGG modules (n = 117) with non-zero copy number in at least one sample, with medians computed in the set of deeply-sequenced healthy (n = 229) and IBD (n = 101) samples. **D)** All biosynthesis modules (n = 88) from the KEGG MODULE database, with medians computed in the set of deeply-sequenced healthy (n = 229) and IBD (n = 101) samples. Where applicable, dashed lines indicate the overall median for all modules, and solid lines connect the points for the same module in each sample group. The IBD sample group is highlighted in red, the NONIBD group in pink, and the HEALTHY group in blue.

Supplementary Figure 6.

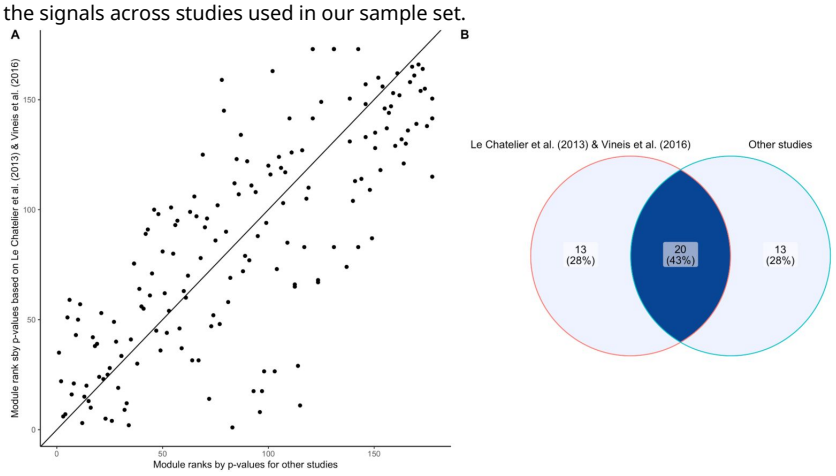
Boxplots of median per-population copy number of 33 IBD-enriched modules for samples from each individual cohort, **A)** with medians computed within each sample (ie, one point per sample) and **B)** with medians computed for each IBD-enriched module (ie, one point per module). The x-axis indicates study of origin. **C)** Boxplots of median per-population copy number of 33 IBD-enriched modules for the 115 samples in the deeply-sequenced set that are not from (Le Chatelier et al. 2013) or (Vineis et al. 2016). The dashed line indicates the overall median for all 33 modules, and solid lines connect the points for the same module in each sample group.



Supplementary Figure 7.

Assessing batch effect of the IBD-enrichment study.

A) Scatter plot comparing the module ranks of Wilcoxon-Mann-Whitney p-values comparing IBD and healthy subjects on (Le Chatelier et al. 2013) and (Vineis et al. 2016) (y axis) and the rest of our dataset (x axis). **B)** Venn diagram displaying the overlap of IBD-enriched modules identified by the 33 smallest p-values in (Le Chatelier et al. 2013) and (Vineis et al. 2016) and the rest of our dataset. There is good agreement (20 out of 33) between the two sets of modules, indicating generalizability of the signals across studies used in our sample set.



References

Aramaki Takuya, Blanc-Mathieu Romain, Endo Hisashi, Ohkubo Koichi, Kanehisa Minoru, Goto Susumu, Ogata Hiroyuki (2019) **KofamKOALA: KEGG Ortholog Assignment Based on Profile HMM and Adaptive Score Threshold** *Bioinformatics* **36**:2251–52

Arkin Adam P., Cottingham Robert W., Henry Christopher S., Harris Nomi L., Stevens Rick L., Maslov Sergei, Dehal Paramvir, et al. (2018) **KBase: The United States Department of Energy Systems Biology Knowledgebase** *Nature Biotechnology* **36**:566–69

Arumugam Manimozhiyan, Raes Jeroen, Pelletier Eric, Le Paslier Denis, Yamada Takuji, Mende Daniel R., Fernandes Gabriel R., et al. (2011) **"Enterotypes of the Human Gut Microbiome."** *Nature* **473**:174–80

Aziz Ramy K., Bartels Daniela, Best Aaron A., DeJongh Matthew, Disz Terrence, Edwards Robert A., Formsma Kevin, et al. (2008) **The RAST Server: Rapid Annotations Using Subsystems Technology** *BMC Genomics* **9**

Belkaid Yasmine, Hand Timothy W. (2014) **Role of the Microbiota in Immunity and Inflammation** *Cell* **157**:121–41

University of Sydney (2016) **EKmeta: metagenome fecal microbiota, Ilumina seq reads of 12 individuals at 2 timepoints** *NCBI BioProject*

Boyle Elizabeth I., Weng Shuai, Gollub Jeremy, Jin Heng, Botstein David, Michael Cherry J., Sherlock Gavin (2004) **GO::TermFinder--Open Source Software for Accessing Gene Ontology Information and Finding Significantly Enriched Gene Ontology Terms Associated with a List of Genes** *Bioinformatics* **20**:3710–15

Buchfink Benjamin, Xie Chao, Huson Daniel H. (2015) **Fast and Sensitive Protein Alignment Using DIAMOND** *Nature Methods* **12**:59–60

Byndloss Mariana X., Bäumler Andreas J. (2018) **The Germ-Organ Theory of Non-Communicable Diseases** *Nature Reviews. Microbiology* **16**:103–10

Cani Patrice D (2018) **Human Gut Microbiome: Hopes, Threats and Promises** *Gut* **67**:1716–25

Capella-Gutiérrez Salvador, Silla-Martínez José M., Gabaldón Toni (2009) **trimAl: A Tool for Automated Alignment Trimming in Large-Scale Phylogenetic Analyses** *Bioinformatics* **25**:1972–73

Chan Patricia P., Lowe Todd M. (2019) **tRNAscan-SE: Searching for tRNA Genes in Genomic Sequences** *Methods in Molecular Biology* **1962**:1–14

Clausen David S., Willis Amy D. (2022) **Evaluating Replicability in Microbiome Data** *Biostatistics* **23**:1099–1114

Coyte Katharine Z., Schluter Jonas, Foster Kevin R. (2015) **The Ecology of the Microbiome: Networks, Competition, and Stability** *Science* **350**:663–66

Degnan Patrick H., Barry Natasha A., Mok Kenny C., Taga Michiko E., Goodman Andrew L. (2014) **Human Gut Microbes Use Multiple Transporters to Distinguish Vitamin B_{12} Analogs and Compete in the Gut** *Cell Host & Microbe* **15**:47–57

Devkota Suzanne, Wang Yunwei, Musch Mark W., Leone Vanessa, Fehlner-Peach Hannah, Nadimpalli Anuradha, Antonopoulos Dionysios A., Jabri Bana, Chang Eugene B. (2012) **Dietary-Fat-Induced Taurocholic Acid Promotes Pathobiont Expansion and Colitis in Il10-/- Mice** *Nature* **487**:104–8

Eddy Sean R (2011) **Accelerated Profile HMM Searches** *PLoS Computational Biology* **7**

Edgar Robert C (2004) **MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput** *Nucleic Acids Research* **32**:1792–97

Eren A. Murat, Vineis Joseph H., Morrison Hilary G., Sogin Mitchell L. (2013) **A Filtering Method to Generate High Quality Short Reads Using Illumina Paired-End Technology** *PloS One* **8**

Fan Yong, Pedersen Oluf (2021) **Gut Microbiota in Human Metabolic Health and Disease** *Nature Reviews. Microbiology* **19**:55–71

Farag Ibrahim F., Biddle Jennifer F., Zhao Rui, Martino Amanda J., House Christopher H., León-Zayas Rosa I. (2020) **Metabolic Potentials of Archaeal Lineages Resolved from Metagenomes of Deep Costa Rica Sediments** *The ISME Journal* **14**:1345–58

Feng Lihui, Raman Arjun S., Hibberd Matthew C., Cheng Jiye, Griffin Nicholas W., Peng Yangqing, Leyn Semen A., Rodionov Dmitry A., Osterman Andrei L., Gordon Jeffrey I. (2020) **Identifying Determinants of Bacterial Fitness in a Model of Human Gut Microbial Succession** *Proceedings of the National Academy of Sciences of the United States of America* **117**:2622–33

Feng Qiang, Liang Suisha, Jia Huijue, Stadlmayr Andreas, Tang Longqing, Lan Zhou, Zhang Dongya, et al. (2015) **Gut Microbiome Development along the Colorectal Adenoma-Carcinoma Sequence** *Nature Communications* **6**

Fithian William, Sun Dennis, Taylor Jonathan (2014) **Optimal Inference After Model Selection** *arXiv*

Franzosa Eric A., Sirota-Madi Alexandra, Avila-Pacheco Julian, Fornelos Nadine, Haiser Henry J., Reinker Stefan, Vatanen Tommi, et al. (2019) **Gut Microbiome Structure and Metabolic Activity in Inflammatory Bowel Disease** *Nature Microbiology* **4**:293–305

Galperin Michael Y., Wolf Yuri I., Makarova Kira S., Alvarez Roberto Vera, Landsman David, Koonin Eugene V. (2021) **COG Database Update: Focus on Microbial Diversity, Model Organisms, and Widespread Pathogens** *Nucleic Acids Research* **49**:D274–81

Geller-McGrath D., Konwar Kishori M., Edgcomb V. P., Pachiadaki M., Roddy J. W., Wheeler T. J., McDermott J. E. (2023) **MetaPathPredict: A Machine Learning-Based Tool for Predicting Metabolic Modules in Incomplete Bacterial Genomes** *bioRxiv* https://doi.org/10.1101/2022.12.21.521254

Gevers Dirk, Kugathasan Subra, Denson Lee A., Vázquez-Baeza Yoshiki, Van Treuren Will, Ren Boyu, Schwager Emma, et al. (2014) **The Treatment-Naive Microbiome in New-Onset Crohn's Disease** *Cell Host & Microbe* **15**:382–92

Goodman Andrew L., McNulty Nathan P., Zhao Yue, Leip Douglas, Mitra Robi D., Lozupone Catherine A., Knight Rob, Gordon Jeffrey I. (2009) **Identifying Genetic Determinants Needed to Establish a Human Gut Symbiont in Its Habitat** *Cell Host & Microbe* **6**:279–89

Halfvarson Jonas, Brislawn Colin J., Lamendella Regina, Vázquez-Baeza Yoshiki, Walters William A., Bramer Lisa M., D'Amato Mauro, et al. (2017) **Dynamics of the Human Gut Microbiome in Inflammatory Bowel Disease** *Nature Microbiology* **2**

Heinken Almut, Hertel Johannes, Thiele Ines (2021) **Metabolic Modelling Reveals Broad Changes in Gut Microbial Metabolism in Inflammatory Bowel Disease Patients with Dysbiosis** *NPJ Systems Biology and Applications* **7**

Henke Matthew T., Kenny Douglas J., Cassilly Chelsi D., Vlamakis Hera, Xavier Ramnik J., Clardy Jon (2019) **Ruminococcus Gnavus, a Member of the Human Gut Microbiome Associated with Crohn's Disease, Produces an Inflammatory Polysaccharide** *Proceedings of the National Academy of Sciences of the United States of America* **116**:12672–77

Hijova E (2019) **Gut Bacterial Metabolites of Indigestible Polysaccharides in Intestinal Fermentation as Mediators of Public Health** *Bratislavske Lekarske Listy* **120**:807–12

Human Microbiome Project Consortium (2012) **A Framework for Human Microbiome Research** *Nature* **486**:215–21

Hunter JD (2007) **Matplotlib: A 2D Graphics Environment** *Computing in Science & Engineering* **9**:90–95

Hyatt Doug, Chen Gwo-Liang, Locascio Philip F., Land Miriam L., Larimer Frank W., Hauser Loren J. (2010) **Prodigal: Prokaryotic Gene Recognition and Translation Initiation Site Identification** *BMC Bioinformatics* **11**

Jaffe Alexander L., Castelle Cindy J., Matheus Carnevali Paula B., Gribaldo Simonetta, Banfield Jillian F. (2020) **The Rise of Diversity in Metabolic Platforms across the Candidate Phyla Radiation** *BMC Biology* **18**

Joossens Marie, Huys Geert, Cnockaert Margo, De Preter Vicky, Verbeke Kristin, Rutgeerts Paul, Vandamme Peter, Vermeire Severine (2011) **Dysbiosis of the Faecal Microbiota in Patients with Crohn's Disease and Their Unaffected Relatives** *Gut* **60**:631–37

Kanehisa Minoru, Furumichi Miho, Sato Yoko, Kawashima Masayuki, Watanabe Mari Ishiguro- (2023) **KEGG for Taxonomy-Based Analysis of Pathways and Genomes** *Nucleic Acids Research* **51**:D587–92

Kanehisa Minoru, Goto Susumu, Sato Yoko, Furumichi Miho, Tanabe Mao (2012) **KEGG for Integration and Interpretation of Large-Scale Molecular Data Sets** *Nucleic Acids Research* **40**:D109–14

Kaplan Gilaad G (2015) **The Global Burden of IBD: From 2015 to 2025** *Nature Reviews. Gastroenterology & Hepatology* **12**:720–27

Karp Peter D., Midford Peter E., Billington Richard, Kothari Anamika, Krummenacker Markus, Latendresse Mario, Ong Wai Kit, et al. (2021) **Pathway Tools Version 23.0 Update: Software for Pathway/genome Informatics and Systems Biology** *Briefings in Bioinformatics* **22**:109–26

Kelly Caleb J., Alexeev Erica E., Farb Linda, Vickery Thad W., Zheng Leon, Eric L Campbell, Kitzenberg David A., et al. (2019) **Oral Vitamin B12 Supplement Is Delivered to the Distal Gut, Altering the Corrinoid Profile and Selectively Depleting Bacteroides in C57BL/6 Mice** *Gut Microbes* **10**:654–62

Khan Israr, Ullah Naeem, Zha Lajia, Bai Yanrui, Khan Ashiq, Zhao Tang, Che Tuanjie, Zhang Chunjiang (2019) **Alteration of Gut Microbiota in Inflammatory Bowel Disease (IBD): Cause or Consequence? IBD Treatment Targeting the Gut Microbiome** *Pathogens* **8** https://doi.org/10.3390/pathogens8030126

Khosravi Arya, Mazmanian Sarkis K. (2013) **Disruption of the Gut Microbiome as a Risk Factor for Microbial Infections** *Current Opinion in Microbiology* **16**:221–27

Knight Rob, Callewaert Chris, Marotz Clarisse, Hyde Embriette R., Debelius Justine W., McDonald Daniel, Sogin Mitchell L. (2017) **The Microbiome and Human Biology** *Annual Review of Genomics and Human Genetics* **18**:65–86

Knox Natalie C., Forbes Jessica D., Van Domselaar Gary, Bernstein Charles N. (2019) **The Gut Microbiome as a Target for IBD Treatment: Are We There Yet?** *Current Treatment Options in Gastroenterology* **17**:115–26

Koek Maud M., Jellema Renger H., van der Greef Jan, Tas Albert C., Hankemeier Thomas (2011) **Quantitative Metabolomics Based on Gas Chromatography Mass Spectrometry: Status and Perspectives** *Metabolomics: Official Journal of the Metabolomic Society* **7**:307–28

Köster Johannes, Rahmann Sven (2012) **Snakemake--a Scalable Bioinformatics Workflow Engine** *Bioinformatics* **28**:2520–22

Kostic Aleksandar D., Xavier Ramnik J., Gevers Dirk (2014) **The Microbiome in Inflammatory Bowel Disease: Current Status and the Future Ahead** *Gastroenterology* **146**:1489–99

Kraus Sarah, Arber Nadir (2009) **Inflammation and Colorectal Cancer** *Current Opinion in Pharmacology* **9**:405–10

Le Chatelier Emmanuelle, Nielsen Trine, Qin Junjie, Prifti Edi, Hildebrand Falk, Falony Gwen, Almeida Mathieu, et al. (7464) **"Richness of Human Gut Microbiome Correlates with Metabolic Markers."** *Nature* **500**:541–46

Lee Michael D (2019) **GToTree: A User-Friendly Workflow for Phylogenomics** *Bioinformatics* **35**:4162–64

Lee Mirae, Chang Eugene B. (2021) **Inflammatory Bowel Diseases (IBD) and the Microbiome—Searching the Crime Scene for Clues** *Gastroenterology* **160**:524–37

Li Dinghua, Liu Chi-Man, Luo Ruibang, Sadakane Kunihiko, Lam Tak-Wah (2015) **"MEGAHIT: An Ultra-Fast Single-Node Solution for Large and Complex Metagenomics Assembly via Succinct de Bruijn Graph."** *Bioinformatics* **31**:1674–76

Lin Yanping, Caldwell Gary W., Li Ying, Lang Wensheng, Masucci John (2020) **Inter-Laboratory Reproducibility of an Untargeted Metabolomics GC-MS Assay for Analysis of Human Plasma** *Scientific Reports* **10**

Lloyd-Price Jason, Arze Cesar, Ananthakrishnan Ashwin N., Schirmer Melanie, Pacheco Julian Avila-, Poon Tiffany W., Andrews Elizabeth, et al. (2019) **Multi-Omics of the Gut Microbial Ecosystem in Inflammatory Bowel Diseases** *Nature* **569**:655–62

Lozupone Catherine A., Stombaugh Jesse, Gonzalez Antonio, Ackermann Gail, Wendel Doug, Vázquez-Baeza Yoshiki, Jansson Janet K., Gordon Jeffrey I., Knight Rob (2013) **Meta-Analyses of Studies of the Human Microbiota** *Genome Research* **23**:1704–14

Machado Daniel, Andrejev Sergej, Tramontano Melanie, Patil Kiran Raosaheb (2018) **Fast Automated Reconstruction of Genome-Scale Metabolic Models for Microbial Species and Communities** *Nucleic Acids Research* **46**:7542–53

Machiels Kathleen, Joossens Marie, Sabino João, De Preter Vicky, Arijs Ingrid, Eeckhaut Venessa, Ballet Vera, et al. (2014) **A Decrease of the Butyrate-Producing Species Roseburia Hominis and Faecalibacterium Prausnitzii Defines Dysbiosis in Patients with Ulcerative Colitis** *Gut* **63**:1275–83

Magnúsdóttir Stefanía, Ravcheev Dmitry, de Crécy-Lagard Valérie, Thiele Ines (2015) **Systematic Genome Assessment of B-Vitamin Biosynthesis Suggests Co-Operation among Gut Microbes** *Frontiers in Genetics* **6**

Marcelino Vanessa R., Welsh Caitlin, Diener Christian, Gulliver Emily L., Rutten Emily L., Young Remy B., Giles Edward M., Gibbons Sean M., Greening Chris, Forster Samuel C. (2023) **Disease-Specific Loss of Microbial Cross-Feeding Interactions in the Human Gut** *bioRxiv* https://doi.org/10.1101/2023.02.17.528570

Martens Eric C., Kelly Amelia G., Tauzin Alexandra S., Brumer Harry (2014) **The Devil Lies in the Details: How Variations in Polysaccharide Fine-Structure Impact the Physiology and Evolution of Gut Microbes** *Journal of Molecular Biology* **426**:3851–65

Maynard Craig L., Elson Charles O., Hatton Robin D., Weaver Casey T. (2012) **Reciprocal Interactions of the Intestinal Microbiota and Immune System** *Nature* **489**:231–41

Minh Bui Quang, Schmidt Heiko A., Chernomor Olga, Schrempf Dominik, Woodhams Michael D., von Haeseler Arndt, Lanfear Robert (2020) **IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era** *Molecular Biology and Evolution* **37**:1530–34

Mistry Jaina, Chuguransky Sara, Williams Lowri, Qureshi Matloob, Salazar Gustavo A., Sonnhammer Erik L. L., Tosatto Silvio C. E., et al. (2021) **Pfam: The Protein Families Database in 2021** *Nucleic Acids Research* **49**:D412–19

Morgan Xochitl C., Tickle Timothy L., Sokol Harry, Gevers Dirk, Devaney Kathryn L., Ward Doyle V., Reyes Joshua A., et al. (2012) **Dysfunction of the Intestinal Microbiome in Inflammatory Bowel Disease and Treatment** *Genome Biology* **13**

Nagalingam Nabeetha A., Lynch Susan V. (2012) **Role of the Microbiota in Inflammatory Bowel Diseases** *Inflammatory Bowel Diseases* **18**:968–84

Nishida Atsushi, Inoue Ryo, Inatomi Osamu, Bamba Shigeki, Naito Yuji, Andoh Akira (2018) **Gut Microbiota in the Pathogenesis of Inflammatory Bowel Disease** *Clinical Journal of Gastroenterology* **11**:1–10

Nitzan Orna, Elias Mazen, Peretz Avi, Saliba Walid (2016) **Role of Antibiotics for Treatment of Inflammatory Bowel Disease** *World Journal of Gastroenterology: WJG* **22**:1078–87

Palleja Albert, Mikkelsen Kristian H., Forslund Sofia K., Kashani Alireza, Allin Kristine H., Nielsen Trine, Hansen Tue H., et al. (2018) **Recovery of Gut Microbiota of Healthy Adults Following Antibiotic Exposure** *Nature Microbiology* **3**:1255–65

Palù Matteo, Basile Arianna, Zampieri Guido, Treu Laura, Rossi Alessandro, Morlino Maria Silvia, Campanaro Stefano (2022) **KEMET – A Python Tool for KEGG Module Evaluation and Microbial Genome Annotation Expansion** *Computational and Structural Biotechnology Journal* https://doi.org/10.1016/j.csbj.2022.03.015

Papa Eliseo, Docktor Michael, Smillie Christopher, Weber Sarah, Preheim Sarah P., Gevers Dirk, Giannoukos Georgia, et al. (2012) **Non-Invasive Mapping of the Gastrointestinal Microbiota Identifies Children with Inflammatory Bowel Disease** *PloS One* **7**

Parks Donovan H., Chuvochina Maria, Chaumeil Pierre-Alain, Rinke Christian, Mussig Aaron J., Hugenholtz Philip (2020) **A Complete Domain-to-Species Taxonomy for Bacteria and Archaea** *Nature Biotechnology* **38**:1079–86

Parks Donovan H., Chuvochina Maria, Rinke Christian, Mussig Aaron J., Chaumeil Pierre-Alain, Hugenholtz Philip (2022) **GTDB: An Ongoing Census of Bacterial and Archaeal Diversity through a Phylogenetically Consistent, Rank Normalized and Complete Genome-Based Taxonomy** *Nucleic Acids Research* **50**:D785–94

Parks Donovan H., Chuvochina Maria, Waite David W., Rinke Christian, Skarshewski Adam, Chaumeil Pierre-Alain, Hugenholtz Philip (2018) **A Standardized Bacterial Taxonomy Based on Genome Phylogeny Substantially Revises the Tree of Life** *Nature Biotechnology* **36**:996–1004

Peng Yu, Leung Henry C. M., Yiu S. M., Chin Francis Y. L. (2012) **"IDBA-UD: A de Novo Assembler for Single-Cell and Metagenomic Sequencing Data with Highly Uneven Depth."** *Bioinformatics* **28**:1420–28

Powell, Elijah J., Leonard Sean P., Kwong Waldan K., Engel Philipp, Moran Nancy A. (2016) **Genome-Wide Screen Identifies Host Colonization Determinants in a Bacterial Gut Symbiont** *Proceedings of the National Academy of Sciences of the United States of America* **113**:13887–92

Prindiville T. P., Sheikh R. A., Cohen S. H., Tang Y. J., Cantrell M. C., Silva Jr J. (2000) **Bacteroides Fragilis Enterotoxin Gene Sequences in Patients with Inflammatory Bowel Disease** *Emerging Infectious Diseases* **6**:171–74

Qin Junjie, Li Yingrui, Cai Zhiming, Li Shenghui, Zhu Jianfeng, Zhang Fan, Liang Suisha, et al. (2012) **A Metagenome-Wide Association Study of Gut Microbiota in Type 2 Diabetes** *Nature* **490**:55–60

Quince Christopher, Ijaz Umer Zeeshan, Loman Nick, Eren Murat A., Saulnier Delphine, Russell Julie, Haig Sarah J., et al. (2015) **Extensive Modulation of the Fecal Metagenome in Children With Crohn's Disease During Exclusive Enteral Nutrition** *American Journal of Gastroenterology* https://doi.org/10.1038/ajg.2015.357

Ramirez Jaime, Guarner Francisco, Fernandez Luis Bustos, Maruy Aldo, Sdepanian Vera Lucia, Cohen Henry (2020) **Antibiotics as Major Disruptors of Gut Microbiota** *Frontiers in Cellular and Infection Microbiology* **10**

Rampelli Simone, Schnorr Stephanie L., Consolandi Clarissa, Turroni Silvia, Severgnini Marco, Peano Clelia, Brigidi Patrizia, Crittenden Alyssa N., Henry Amanda G., Candela Marco (2015) **Metagenome Sequencing of the Hadza Hunter-Gatherer Gut Microbiota** *Current Biology: CB* **25**:1682–93

Raymond Frédéric, Ouameur Amin A., Déraspe Maxime, Iqbal Naeem, Gingras Hélène, Dridi Bédis, Leprohon Philippe, et al. (2016) **The Initial State of the Human Gut Microbiome Determines Its Reshaping by Antibiotics** *The ISME Journal* **10**:707–20

Reiter Taylor E., Irber Luiz, Gingrich Alicia A., Haynes Dylan, Tessa Pierce-Ward N., Brooks Phillip T., Mizutani Yosuke, et al. (2022) **Meta-Analysis of Metagenomes via Machine Learning and Assembly Graphs Reveals Strain Switches in Crohn's Disease** *bioRxiv* https://doi.org/10.1101/2022.06.30.498290

Rhodes Jonathan M (2007) **"The Role of Escherichia Coli in Inflammatory Bowel Disease** *Gut*

Saitoh Shin, Noda Satoshi, Aiba Yuji, Takagi Atsushi, Sakamoto Mitsuo, Benno Yoshimi, Koga Yasuhiro (2002) **"Bacteroides Ovatus as the Predominant Commensal Intestinal Microbe Causing a Systemic Antibody Response in Inflammatory Bowel Disease."** *Clinical and Diagnostic Laboratory Immunology* **9**:54–59

Sartor R. Balfour (2006) **Mechanisms of Disease: Pathogenesis of Crohn's Disease and Ulcerative Colitis** *Nature Clinical Practice. Gastroenterology & Hepatology* **3**:390–407

Schirmer Melanie, Franzosa Eric A., Lloyd-Price Jason, McIver Lauren J., Schwager Randall, Poon Tiffany W., Ananthakrishnan Ashwin N., et al. (2018) **Dynamics of Metatranscription in the Inflammatory Bowel Disease Gut Microbiome** *Nature Microbiology* **3**:337–46

Schirmer Melanie, Garner Ashley, Vlamakis Hera, Xavier Ramnik J. (2019) **Microbial Genes and Pathways in Inflammatory Bowel Disease** *Nature Reviews. Microbiology* **17**:497–511

Seemann Torsten (2022) **Barrnap: Bacterial Ribosomal RNA Predictor. Github. Accessed September 30**

Shaffer Michael, Borton Mikayla A., McGivern Bridget B., Zayed Ahmed A., La Rosa Sabina Leanti, Solden Lindsey M., Liu Pengfei, et al. (2020) **DRAM for Distilling Microbial Metabolism to Automate the Curation of Microbiome Function** *Nucleic Acids Research* **48**:8883–8900

Shaiber Alon, Willis Amy D., Delmont Tom O., Roux Simon, Chen Lin-Xing, Schmid Abigail C., Yousef Mahmoud, et al. (2020) **Functional and Genetic Markers of Niche Partitioning among Enigmatic Members of the Human Oral Microbiome** *Genome Biology* **21**

Shan Yue, Lee Mirae, Chang Eugene B. (2022) **The Gut Microbiome and Inflammatory Bowel Diseases** *Annual Review of Medicine* **73**:455–68

Sinha Rashmi, Abu-Ali Galeb, Vogtmann Emily, Fodor Anthony A., Ren Boyu, Amir Amnon, Schwager Emma, et al. (2017) **Assessment of Variation in Microbial Community Amplicon Sequencing by the Microbiome Quality Control (MBQC) Project Consortium** *Nature Biotechnology* **35**:1077–86

Sorbara Matthew T., Pamer Eric G. (2022) **Microbiome-Based Therapeutics** *Nature Reviews. Microbiology* **20**:365–80

Steinegger Martin, Söding Johannes (2017) **MMseqs2 Enables Sensitive Protein Sequence Searching for the Analysis of Massive Data Sets** *Nature Biotechnology* **35**:1026–28

Turnbaugh Peter J., Hamady Micah, Yatsunenko Tanya, Cantarel Brandi L., Duncan Alexis, Ley Ruth E., Sogin Mitchell L., et al. (2009) **A Core Gut Microbiome in Obese and Lean Twins** *Nature* **457**:480–84

Vineis J. H., Ringus D. L., Morrison H. G., Delmont T. O., Dalal S., Raffals L. H., Antonopoulos D. A., et al. (2016) **Patient-Specific Bacteroides Genome Variants in Pouchitis** *mBio* **7**

Watson Andrea R., Füssel Jessika, Veseli Iva, DeLongchamp Johanna Zaal, Silva Marisela, Trigodet Florian, Lolans Karen, et al. (2023) **Metabolic Independence Drives Gut Microbial Colonization and Resilience in Health and Disease** *Genome Biology* **24**

Wen Chengping, Zheng Zhijun, Shao Tiejuan, Liu Lin, Xie Zhijun, Le Chatelier Emmanuelle, He Zhixing, et al. (2017) **Quantitative Metagenomics Reveals Unique Gut Microbiome Biomarkers in Ankylosing Spondylitis** *Genome Biology* **18**

Wickham Hadley (2016) **ggplot2: Elegant Graphics for Data Analysis** New York: Springer-Verlag

Woting Anni, Blaut Michael (2016) **The Intestinal Microbiota in Metabolic Disease** *Nutrients* **8**

Xie Hailiang, Guo Ruijin, Zhong Huanzi, Feng Qiang, Lan Zhou, Qin Bingcai, Ward Kirsten J., et al. (2016) **Shotgun Metagenomics of 250 Adult Twins Reveals Genetic and Environmental Impacts on the Gut Microbiome** *Cell Systems* **3**:572–84

Ye Yuzhen, Doak Thomas G. (2009) **A Parsimony Approach to Biological Pathway Reconstruction/inference for Genomes and Metagenomes** *PLoS Computational Biology* **5**

Zhou Zhichao, Tran Patricia Q., Breister Adam M., Liu Yang, Kieft Kristopher, Cowley Elise S., Karaoz Ulas, Anantharaman Karthik (2021) **METABOLIC: High-Throughput Profiling of Microbial Genomes for Functional Traits, Metabolism, Biogeochemistry, and Community-Scale Functional Networks** *Research Square* https://doi.org/10.21203/rs.3.rs-965097/v1

Zimmermann Johannes, Kaleta Christoph, Waschina Silvio (2021) **Gapseq: Informed Prediction of Bacterial Metabolic Pathways and Reconstruction of Accurate Metabolic Models** *Genome Biology* **22**

Zimmermann Michael, Zimmermann-Kogadeeva Maria, Wegmann Rebekka, Goodman Andrew L. (2019) **Mapping Human Microbiome Drug Metabolism by Gut Bacteria and Their Genes** *Nature* **570**:462–67

Zong Xin, Fu Jie, Xu Bocheng, Wang Yizhen, Jin Mingliang (2020) **Interplay between Gut Microbiota and Antimicrobial Peptides** *Animal Nutrition (Zhongguo Xu Mu Shou Yi Xue Hui* **6**:389–96

Zorrilla Francisco, Buric Filip, Patil Kiran R., Zelezniak Aleksej (2021) **metaGEM: Reconstruction of Genome Scale Metabolic Models Directly from Metagenomes** *Nucleic Acids Research* **49**

Boyle Elizabeth I, Weng Shuai, Gollub Jeremy, Jin Heng, Botstein David, Michael Cherry J, Sherlock Gavin (2004) **"GO::TermFinder–open Source Software for Accessing Gene Ontology Information and Finding Significantly Enriched Gene Ontology Terms Associated with a List of Genes." Bioinformatics (Oxford** *England* **20**:3710–15 https://doi.org/10.1093/bioinformatics/bth456

Author information

Iva Veseli
Biophysical Sciences Program, The University of Chicago, Chicago, USA, Department of Medicine, The University of Chicago, Chicago, USA
ORCID iD: 0000-0003-2390-5286

For correspondence: iveseli@uchicago.edu

Yiqun T Chen
Data Science Institute and Department of Biomedical Data Science, Stanford University, Stanford, USA

Matthew S Schechter
Department of Medicine, The University of Chicago, Chicago, USA, Committee on Microbiology, The University of Chicago, Chicago, USA
ORCID iD: 0000-0002-8435-3203

Chiara Vanni
MARUM Center for Marine Environmental Sciences, University of Bremen, Bremen, Germany
ORCID iD: 0000-0002-1124-1147

Emily C Fogarty
Department of Medicine, The University of Chicago, Chicago, USA, Committee on Microbiology, The University of Chicago, Chicago, USA
ORCID iD: 0000-0002-8957-9922

Andrea R Watson
Department of Medicine, The University of Chicago, Chicago, USA, Committee on Microbiology, The University of Chicago, Chicago, USA
ORCID iD: 0000-0003-0128-6795

Bana Jabri
Department of Medicine, The University of Chicago, Chicago, USA
ORCID iD: 0000-0001-7427-4424

Ran Blekhman
Department of Medicine, The University of Chicago, Chicago, USA
ORCID iD: 0000-0003-3218-613X

Amy D Willis
Department of Biostatistics, University of Washington, Seattle, USA
ORCID iD: 0000-0002-2802-4317

Michael K Yu
Toyota Technological Institute at Chicago, Chicago, USA

Antonio Fernàndez-Guerra
Lundbeck Foundation GeoGenetics Centre, GLOBE Institute, University of Copenhagen, Copenhagen, Denmark
ORCID iD: 0000-0002-8679-490X

Jessika Füssel
Department of Medicine, The University of Chicago, Chicago, USA, Institute for Chemistry and Biology of the Marine Environment, University of Oldenburg, Oldenburg, Germany
ORCID iD: 0000-0002-4210-2318

For correspondence: jessika.fuessel@uni-oldenburg.de

A Murat Eren
Department of Medicine, The University of Chicago, Chicago, USA, Institute for Chemistry and Biology of the Marine Environment, University of Oldenburg, Oldenburg, Germany, Marine 'Omics Bridging Group, Max Planck Institute for Marine Microbiology, Bremen, Germany, Alfred Wegener Institute for Polar and Marine Research, Bremerhaven, Germany, Helmholtz Institute for Functional Marine Biodiversity, Oldenburg, Germany
ORCID iD: 0000-0001-9013-4827

For correspondence: meren@hifmb.de

Editors

Reviewing Editor
Peter Turnbaugh
University of California, San Francisco, San Francisco, United States of America

Senior Editor
Wendy Garrett
Harvard T.H. Chan School of Public Health, Boston, United States of America

Reviewer #1 (Public review):

In this work, Veseli et al. present a computational framework to infer the functional diversity of microbiomes in relation to microbial diversity directly from metagenomic data. The framework reconstructs metabolic modules form metagenomes and calculates the per-population copy number of each module, resulting in the proportion of microbes in the sample carrying certain genes. They applied this framework to a dataset of gut microbiomes from 109 inflammatory bowel disease (IBD) patients, 78 patients with other gastrointestinal conditions, and 229 healthy controls. The found that the microbiomes of IBD patients were enriched in a high fraction of metabolic pathways, including biosynthesis pathways such as those for amino acids, vitamins, nucleotides, and lipids. Hence, they had higher metabolic independence compared with healthy controls. To an extent, the authors also found a pathway enrichment suggesting higher metabolic independence in patients with gastrointestinal conditions other than IBD indicating this could be a signal for a general loss in host health. Finally, a machine learning classifier using high metabolic independence in microbiomes could predict IBD with good accuracy. Overall, this is an interesting and well-

written article and presents a novel workflow that enables a comprehensive characterization of microbiome cohorts.

Comments on revisions:

I believe that after the second round of revisions, the Reviewers sufficiently addressed the comments and improved the manuscript. Open questions have been answered. I have no further comments.

https://doi.org/10.7554/eLife.89862.2.sa3

Reviewer #2 (Public review):

This study builds upon the team's recent discovery that antibiotic treatment and other disturbances favours the persistence of bacteria with genomes that encode complete modules for the synthesis of essential metabolites (Watson et al. 2023). Veseli and collaborators now provide an in-depth analysis of metabolic pathway completeness within microbiomes, finding strong evidence for an enrichment of bacteria with high metabolic independence in the microbiomes associated with IBD and other gastrointestinal disorders. Importantly, this study provides a new open-source software to facilitate the reconstruction of metabolic pathways, estimate their completeness and normalize their results according to species diversity. Finally, this study also shows that metabolic independence of microbial communities can be used as a marker of dysbiosis. The function-based health index proposed here is more robust to individual's lifestyles and geographic origin than previously proposed methods based on bacterial taxonomy.

The implications of this study have the potential to spur a paradigm shift in the field. It shows that certain bacterial taxa that have been consistently associated with disease might not be harmful to their host as previously thought. These bacteria seem to be the only species that are able to survive in a stressed gut environment. They might even be important to rebuild a healthy microbiome (although the authors are careful in not making this speculation).

This paper provides an in-depth discussion of the results, and limitations are clearly addressed throughout the manuscript (see also the supplementary files for an in-depth assessment of the robustness of the methods). Some of the potential limitations relate to the use of large publicly available datasets, where sample processing and the definition of healthy status varies between studies. The authors have recognised these issues and their results were robust to analyses performed at a per-cohort basis. The potential limitations therefore are unlikely to have affected the conclusions of this study.

Overall, this is manuscript is a magnificent contribution to the field, likely to inspire many other studies to come.

Comments on revisions:

The authors have performed a detailed assessment of the accuracy and robustness of their new methods, and included an informative session comparing their new approach with existing ones. The new analyses have strengthened the manuscript, and the results support the biological interpretations of the study.
I commend the authors for the effort and the excellent research.

https://doi.org/10.7554/eLife.89862.2.sa2

Reviewer #3 (Public review):

The major strength of this manuscript is the "anvi-estimate-metabolism' tool, which is already accessible online, extensively documented, and potentially broadly useful to microbial ecologists. Inclusion of extensive benchmarking and validation on simulated metagenomes has further increased confidence in this approach. Further, the conceptual insights raise interesting hypotheses that could be pursued in follow-on experimental work.

Comments on revisions:

Thank you for the very thorough response and congratulations!

https://doi.org/10.7554/eLife.89862.2.sa1

Author response:

The following is the authors' response to the original reviews.

> ***Response to Public Reviewer Comments:***
>
> ***Reviewer 1:***
>
> *In this work, Veseli et al. present a computational framework to infer the functional diversity of microbiomes in relation to microbial diversity directly from metagenomic data. The framework reconstructs metabolic modules from metagenomes and calculates the per-population copy number of each module, resulting in the proportion of microbes in the sample carrying certain genes. They applied this framework to a dataset of gut microbiomes from 109 inflammatory bowel disease (IBD) patients, 78 patients with other gastrointestinal conditions, and 229 healthy controls. They found that the microbiomes of IBD patients were enriched in a high fraction of metabolic pathways, including biosynthesis pathways such as those for amino acids, vitamins, nucleotides, and lipids. Hence, they had higher metabolic independence compared with healthy controls. To an extent, the authors also found a pathway enrichment suggesting higher metabolic independence in patients with gastrointestinal conditions other than IBD indicating this could be a signal for a general loss in host health. Finally, a machine learning classifier using high metabolic independence in microbiomes could predict IBD with good accuracy. Overall, this is an interesting and well-written article and presents a novel workflow that enables a comprehensive characterization of microbiome cohorts.*

We thank the reviewer for their interest in our study, their summary of its findings, and their kind words about the manuscript quality.

> ***Reviewer 2:***
>
> *This study builds upon the team's recent discovery that antibiotic treatment and other disturbances favour the persistence of bacteria with genomes that encode complete modules for the synthesis of essential metabolites (Watson et al. 2023). Veseli and collaborators now provide an in-depth analysis of metabolic pathway completeness within microbiomes, finding strong evidence for an enrichment of bacteria with high metabolic independence in the microbiomes associated with IBD and other gastrointestinal disorders. Importantly, this study provides new open-source software to facilitate the reconstruction of metabolic pathways, estimate their completeness and normalize their results according to species diversity. Finally, this study also shows that the metabolic independence of microbial communities can be used as a marker of dysbiosis. The function-based health index proposed here is more robust to individuals'*

> *lifestyles and geographic origin than previously proposed methods based on bacterial taxonomy.*
>
> *The implications of this study have the potential to spur a paradigm shift in the field. It shows that certain bacterial taxa that have been consistently associated with disease might not be harmful to their host as previously thought. These bacteria seem to be the only species that are able to survive in a stressed gut environment. They might even be important to rebuild a healthy microbiome (although the authors are careful not to make this speculation).*
>
> *This paper provides an in-depth discussion of the results, and limitations are clearly addressed throughout the manuscript. Some of the potential limitations relate to the use of large publicly available datasets, where sample processing and the definition of healthy status varies between studies. The authors have recognised these issues and their results were robust to analyses performed on a per-cohort basis. These potential limitations, therefore, are unlikely to have affected the conclusions of this study.*
>
> *Overall, this manuscript is a magnificent contribution to the field, likely to inspire many other studies to come.*

We thank the reviewer for their endorsement of our study and their precision regarding the evaluation of its strengths. We also appreciate their high expectations for its impact in the field.

> ***Reviewer 3:***
>
> *The major strength of this manuscript is the "anvi-estimate-metabolism' tool, which is already accessible online, extensively documented, and potentially broadly useful to microbial ecologists.*

We thank the reviewer for their recognition of the computational advances in this study. We also thank the reviewer for their suggestions that we have addressed below, which allowed us to strengthen our manuscript.

> *However, the context for this tool and its validation is lacking in the current version of the manuscript. It is unclear whether similar tools exist; if so, it would help to benchmark this new tool against prior methods.*

The reviewer brings up a very good point about the lack of context for the ànvi-estimate-metabolism ùprogram. While our efforts that led to the emergence of this software included detailed benchmarking efforts, a formal assessment of its performance and accuracy was indeed lacking. We are thankful for our reviewer to point this out, which motivated us to perform additional analyses to address such concerns. Our revision contains a new, 34-page long supplementary information file (Supplementary File 2) that includes a section titled “Comparison of anvi-estimate-metabolism to existing tools for metabolism reconstruction”. The text therein describes the landscape of currently available software for metabolism reconstruction and describes the features that make ànvi-estimate-metabolism ùnique – namely, (1) its implementation of metrics that make it suitable for metagenome-level analyses (i.e., pathway copy number and stepwise interpretation of pathway definitions) and (2) its ability to process user-defined metabolic pathways rather than exclusively relying on KEGG. As described in that section, there is currently no other tool that can compute copy numbers of metabolic pathways from metagenomic data. Hence, it is not quite possible to benchmark the copy number methodology used in our study against prior methods; however, our benchmarking of this functionality with synthetic genomes and metagenomes (described

later in this document) does provide necessary quantitative insights into its accuracy and efficiency.

While comparison of the copy number calculations to other tools was not possible due to the unique nature of this functionality, it was possible to benchmark our gene function annotation methodology against existing tools that also annotate genes with KEGG KOfams, which is a step commonly used by various tools that aim to estimate metabolic potential in genomes and metagenomes. In the anvi'o software ecosystem the annotation of genes for metabolic reconstruction is implemented in ànvi-run-kegg-kofams,̀ and represents a step that is required by ànvi-estimate-metabolism.̀ As our comparisons were quite extensive and involved additional researchers, we described them in another study which we titled "Adaptive adjustment of significance thresholds produces large gains in microbial gene annotations and metabolic insights" (doi:10.1101/2024.07.03.601779) that is now cited from within our revision in the appropriate context. Briefly, our comparison of anvi'o, Kofamscan, and MicrobeAnnotator using 396 publicly-available bacterial genomes from 11 families demonstrated that ànvi-run-kegg-kofams ìs able to identify an average of 12.8% more KO annotations per genome than the other tools, especially in families commonly found in the gut environment (Figure 1). Furthermore, anvi'o recovered the highest proportion of annotations that were independently validated using eggNOG-mapper. Our comparisons also showed that annotations from anvi'o yield at least 11.6% more complete metabolic modules than Kofamscan or MicrobeAnnotator, including the identification of butyrate biosynthesis in *Lachnospiraceae* genomes at rates similar to manual identification of this pathway in this clade (Figure 2a). Overall, our findings that are now described extensively in DOI:10.1101/2024.07.03.601779 show that our method captures high-quality annotations for accurate downstream metabolism estimates.

We hope these new data help increase the reviewer's confidence in our results.

> *Simulated datasets could be used to validate the approach and test its robustness to different levels of bacterial richness, genome sizes, and annotation level.*

We thank the reviewer for this suggestion. It was an extremely useful exercise that not only helped us elucidate the nuances of our approach, but also enabled us to further highlight its strengths in our manuscript. We created simulated datasets including a total of 409 synthetic metagenomes that we used to test the robustness of our approach to different genome sizes, community sizes, and levels of diversity. Overall, our tests with these synthetic metagenomes demonstrated that our approach of computing PPCN values to summarize the metabolic capacity within a metagenomic community is accurate and robust to differences in all three critical variables. Most of these variables were weakly correlated between PPCN or PPCN accuracy, and the few correlations that were stronger in fact further supported our original hypothesis that we generated from our comparisons of healthy and IBD gut metagenomes. The methods and results of our validation efforts are explained in detail in our new Supplementary File 2 (see the section titled "Validation of per-population copy number (PPCN) approach on simulated metagenomic data"), but we copy here the subsection that summarizes our findings for the reviewer's convenience:

Overall impact on the comparison between healthy and IBD gut metagenomes

"In summary, our validation strategy revealed good accuracy at estimating metagenome-level metabolic capacity relative to our genome-level knowledge in the simulated data. While it often underestimated average genomic completeness by ignoring partial copies of metabolic pathways and often overestimated average genomic copy number due to the effect of pathway complementarity between different community members, the magnitude of error was overall limited in range and the error distributions were centered at or near 0. Furthermore, we observed these broad error trends in all cases we tested, and therefore we

expect that they would also apply to both sample groups in our comparative analysis. Thus, we next considered how the PPCN approach might have influenced our analyses that considered metagenomes from healthy individuals and from those who have IBD – two groups that differed from one another with respect to some of the variables considered in our tests.

Most of the correlations between PPCN or PPCN accuracy and sample parameters were weak, yet significant (Table 1). They showed that community size and diversity level have limited influence on the PPCN calculation, while genome size does not influence its accuracy. The only exception was the moderate correlation between PPCN and genome size, particularly for the subset of IBD-enriched pathways. It was a negative correlation with the proportion of small genomes in a metagenome, indicating that PPCN values for these pathways are larger when there are more large genomes in the community and suggesting that these pathways tend to occur frequently in larger genomes. This is in line with our observation that IBD communities contain more large genomes and therefore confirms our interpretation that the populations surviving in the IBD gut microbiome are those with the genomic space to encode more metabolic capacities.

If we consider even the weak correlations, two of those relationships indicate that our approach would be more accurate for IBD metagenomes than for healthy metagenomes. For instance, PPCN accuracy was slightly higher for smaller communities (as in IBD samples), with a weakly positive correlation between PPCN error and community size. It was also slightly more accurate for less diverse communities (as in IBD samples), with a weakly positive correlation between PPCN error and number of phyla. The only opposing trend was the weakly positive correlation between PPCN error and proportion of smaller genomes, which favors higher accuracy in communities with smaller genomes (as in healthy samples). Given that our analysis focuses on the pathways enriched in IBD samples, an overall higher accuracy in IBD samples would increase the confidence in our enrichment results.

We also examined the accuracy of our method to predict the number of populations within a metagenome based on the distribution and frequency of single-copy core genes (i.e., the denominator in the calculation of PPCN). Our benchmarks show that the estimates are overall accurate, where most errors reflect a negligible amount of underestimations of the actual number of populations. Errors occurred more frequently for the realistic synthetic assemblies generated from simulated short read data than for the ideal synthetic assemblies generated from the combination of genomic contigs. The correlations between estimation accuracy and sample parameters indicated that the population estimates are more accurate for smaller communities and communities with more large genomes, as in IBD samples (Table 2). Thus, this method is more likely to underestimate the community size in healthy samples, and these errors could lead to overestimation of PPCN in healthy samples relative to IBD samples. Thus, the enrichment of a given pathway in the IBD samples would have to overcome its relative overestimation in the healthy sample group, making it more likely that we identified pathways that were truly enriched in the IBD communities.

Overall, the consideration of our simulations in the context of healthy vs IBD metagenomes suggest that slight biases in our estimates as a function of unequal diversity with sample groups should have driven PPCN calculations towards a conclusion that is opposite of our observations under neutral conditions. Thus, clear differences between healthy vs IBD metagenomes that overcome these biases suggest that biology, and not potential bioinformatics artifacts, is the primary driver of our observations."

Accordingly, we have added the following sentence summarizing the validation results to our paper:

"Our validation of this method on simulated metagenomic data demonstrated that it is accurate in capturing metagenome-level metabolic capacity relative to genome-level

metabolic capacity estimated from the same data (Supplementary File 2, Supplementary Table 6)."

Early in this process of validation, we identified and fixed two minor bugs in our codebase. The bugs did not affect the results of our paper and therefore did not warrant a re-analysis of our data. The first bug, which is detailed in the Github issue https://github.com/merenlab/anvio/issues/2231 and fixed in the pull request https://github.com/merenlab/anvio/pull/2235, led to the overestimation of the number of microbial populations in a metagenome when the metagenome contains both Bacteria and Archaea. None of the gut metagenomes analyzed in our paper contained archaeal populations, so this bug did not affect our community size estimates.

The second bug, which is detailed in the Github issue https://github.com/merenlab/anvio/issues/2217 and fixed in the pull request https://github.com/merenlab/anvio/pull/2218, caused inflation of stepwise copy numbers for a specific type of metabolic pathway in which the definition contained an inner parenthetical clause. This bug affected only 3 pathways in the KEGG MODULE database we used for our analysis, M00083, M00144, and M00149. It is worth noting that one of those pathways, M00083, was identified as an IBD-enriched module in our analysis. However, the copy number inflation resulting from this bug would have occurred equivalently in both the healthy and IBD sample groups and thus should not have impacted our comparative analysis.

Regardless, we are grateful for the suggestion to validate our approach since it enabled us to identify and eliminate these minor issues.

> *The concept of metabolic independence was intriguing, although it also raises some concerns about the overinterpretation of metagenomic data. As mentioned by the authors, IBD is associated with taxonomic shifts that could confound the copy number estimates that are the primary focus of this analysis. It is unclear if the current results can be explained by IBD-associated shifts in taxonomic composition and/or average genome size. The level of prior knowledge varies a lot between taxa; especially for the IBD-associated gamma-Proteobacteria.*

The reviewer brings up an important point, and we are thankful for the opportunity to clarify the impact of taxonomy on our analysis. Though IBD has been associated with taxonomic shifts in the gut microbiome, a major problem with such associations is that the taxonomic signal is extremely variable, leading to inconsistency in the observed shifts across different studies (doi:https://doi.org/10.3390/pathogens8030126). Indeed, one of the most comprehensive prior studies into this topic demonstrated that inter-individual variation is the largest contributor to all multi-omic measurements aiming to differentiate between the gut microbiome of individuals with IBD from that of healthy individuals, including taxonomy (doi:10.1038/s41586-019-1237-9). We therefore took a different approach to study this question that is independent of taxonomy, by focusing on metabolic potential estimated directly from metagenomes to elucidate an ecological explanation behind the reduced diversity of the IBD gut microbiome, which studies of taxonomic composition alone are not able to provide. Furthermore, the variability inherent to taxonomic profiles of the gut microbiome makes it unlikely that taxonomic shifts could confound our analysis, especially given our large sample set encompassing a variety of individuals with different origins, ages, and genders.

We agree with the reviewer that our level of prior knowledge varies substantially across taxa. Regardless, the only prior knowledge with any bearing on our ability to estimate metabolic capacity in a taxonomy-independent manner is the extent of sequence diversity captured by our annotation models for the enzymes used in metabolic pathways. During our analysis, we had observed that metagenomes in the healthy group had fewer gene annotations than those

in the IBD group and we therefore shared the reviewer's concern about potential annotation bias, whereby less-studied genomes are not always incorporated into the Hidden Markov Models for annotating KEGG Orthologs, perhaps making it more likely for us to miss annotations in these genomes (and leading to lower completeness scores for metabolic pathways in the healthy samples). Our annotation method partially addresses this limitation by taking a second look at any unannotated genes and mindfully relaxing the bit score similarity thresholds to capture annotations for any genes that are slightly too different from reference sequences for annotation with default thresholds. As mentioned previously, our recent preprint demonstrates the efficacy of this strategy (doi:10.1101/2024.07.03.601779). To further address this concern, we also investigated the extent of distant homology in these metagenomes using AGNOSTOS (doi:https://doi.org/10.7554/eLife.67667), which showed a higher proportion of unknown genes in the healthy metagenomes and suggested that a substantial portion of the unannotated genes are not distant homologs of known enzymes that we failed to annotate due to lack of prior knowledge about them, but rather are completely novel functions. To describe these results, we added the following paragraph and two accompanying figures (Supplementary Figure 4g-h) to the section "Differential annotation efficiency between IBD and Healthy samples" in Supplementary File 1:

"To understand the potential origins of the reduced annotation rate in healthy metagenomes, we ran AGNOSTOS (Vanni et al. 2022) to classify known and unknown genes within the healthy and IBD sample groups. AGNOSTOS clusters genes to contextualize them within an extensive reference dataset and then categorizes each gene as 'known' (has homology to genes annotated with Pfam domains of known function), 'genomic unknown' (has homology to genes in genomic reference databases that do not have known functional domains), or 'environmental unknown' (has homology to genes from metagenomes or MAGs that do not have known functional domains). The resulting classifications confirm that healthy metagenomes contain fewer 'known' genes than metagenomes in the IBD sample group – the proportion of 'known' genes classified by AGNOSTOS is about 3.0% less in the healthy metagenomes than in the IBD sample group, which is similar to the ~3.5% decrease in the proportion of 'unannotated' genes observed by simply counting the number of genes with at least one functional annotation (Supplementary Figure 4g-h, Supplementary Table 1e). Furthermore, the majority of the unannotated genes in either sample group were categorized by AGNOSTOS as 'genomic unknown' (Supplementary Figure 4g), suggesting that the unannotated sequences are genes without biochemically-characterized functions currently associated with them and are thus legitimately lacking a functional annotation in our analysis, rather than representing distant homologs of known protein families that we failed to annotate. Based upon the classifications, a systematic technical bias is unlikely driving the annotation discrepancy between the sample groups."

Furthermore, we have already discussed this limitation and its implications in our manuscript (see section "Key biosynthetic pathways are enriched in microbial populations from IBD samples"). To further clarify that our approach is independent of taxonomy, we have now also amended the following statement in our introduction:

"Here we implemented a high-throughput, taxonomy-independent strategy to estimate metabolic capabilities of microbial communities directly from metagenomes and investigate whether the enrichment of populations with high metabolic independence predicts IBD in the human gut."

Finally, the reviewer is also correct that genome size is a part of the equation, as genome size and level of metabolic capacity are inextricable. In fact, we observed this in our analysis, as already stated in our paper:

"HMI genomes were on average substantially larger (3.8 Mbp) than non-HMI genomes (2.9 Mbp) and encoded more genes (3,634 vs. 2,683 genes, respectively)"

Since larger genomes have the space to encode more functional capacity, it follows that having higher metabolic independence would require a microbe to have a larger genome. The validation of our method on simulated metagenomic data supported this idea by demonstrating that the IBD-enriched metabolic pathways are commonly identified in large genomes. The validation also proved that genome size does not influence the accuracy of our approach (Supplementary File 2).

> *It can be difficult to distinguish genes for biosynthesis and catabolism just from the KEGG module names and the new normalization tool proposed herein markedly affects the results relative to more traditional analyses.*

We agree with the reviewer that KEGG module names do not clearly indicate the presence of biosynthetic genes of interest. That said, KEGG is a commonly-used and extensively-curated resource, and many biologists (including ourselves) trust their categorization of genes into pathways. We hope that readers who are interested in specific genes within our results would make use of our publicly-available datasets (which include gene annotations) to conduct a targeted analysis based on their expertise and research question.

However, we would like to respectfully note that the ability to distinguish the genes within each KEGG module may not be very useful to most readers, and is unlikely to have a meaningful impact in our findings. As the reviewer most likely appreciates, the presence of individual genes in isolation can be insufficient to indicate biosynthetic capacity, considering that 1) most biosynthetic pathways involve several biochemical conversions requiring a series of enzymes, 2) enzymes are often multi-functional rather than exclusive to one pathway, and 3) different organisms in a community may utilize enzymes encoded by different genes to perform the same or similar biochemical reaction in a pathway. We therefore made the choice to analyze metabolic capacity at the pathway level, because this would better reflect the biosynthetic abilities encoded by the multiple microbial populations within each metagenome.

The reviewer also suggests that our novel normalization method affects our results, yet we believe that this normalization strategy is one of the strengths of our study in comparison to 'more traditional analyses' as it enables an appropriate comparison between metagenomes describing microbial communities of dramatically different degrees of richness. Indeed, we suspect that the lack of normalization in more traditional analyses may be one reason why prior analyses have so far failed to uncover any mechanistic explanation for the loss of diversity in the IBD gut microbiome. We hope that our validation efforts were sufficiently convincing in demonstrating the suitability of our approach, and copy here a particularly illuminating section of the validation results that we have added to Supplementary Information File 2:

"As expected, we observed a significant positive correlation between metagenomic copy number (the numerator of PPCN) and community size in each group, likely driven by the increase in the copy number of core metabolic pathways in larger communities (Supplementary Figure 18). Interestingly, this correlation was much stronger for the subset of IBD-enriched pathways ($0.49 <= R <= 0.67$) than for all modules ($0.12 <= R <=0.13$).

"However, the correlation was much weaker and often nonsignificant for the normalized PPCN data in both groups of modules (all modules: $0.01 < R < 0.04$, enriched modules: $0.04 < R < 0.09$, Supplementary Table 6b, Supplementary Figure 19), which demonstrates the suitability of our normalization method to remove the effect of community size in comparisons of metagenome-level metabolic capacity."

> *As such, it seems safer to view the current analysis as hypothesis-generating, requiring additional data to assess the degree to which metabolic dependencies are linked to IBD.*

We certainly agree with the reviewer that our study, similar to the vast majority of studies published every year, is a hypothesis-generating work. Any idea proposed in any scientific study in life sciences will certainly benefit from additional data analyses, and therefore we respectfully do not accept this as a valid criticism of our work. The inception of this study is linked to an earlier work that hypothesized high metabolic independence as a determinant of microbial fitness in stressed gut communities (doi:10.1186/s13059-023-02924-x), which lacked validation on larger sets of data. Our study tests this original hypothesis using a large number of metagenomes, and lends further support for it with approaches that are now better validated. Furthermore, there are other studies that agree with our interpretation of the data (doi:10.1101/2023.02.17.528570, doi:10.1038/s41540-021-00178-6), and we look forward to more computational and/or experimental work in the future to generate more evidence to evaluate these insights further.

> ***Response to Recommendations for the Authors***
>
> ***Reviewer 1:***
>
> *My main comments include:*
>
> *- From the results reported in lines 178-185, it seems that metabolic pathways in general were enriched in IBD microbiomes, not specifically biosynthetic pathways. Can we really say then that the signal is specific for biosynthesis capabilities?*

We apologize for the confusion here. When we read the text again, we ourselves were confused with our phrasing.

The reviewer is correct that a similar proportion of both biosynthetic and non-biosynthetic pathways had elevated per-population copy number (PPCN) values in the IBD samples. However, the low microbial diversity associated with IBD and the on average larger genome size of individual populations contributes to this relative enrichment of the majority of metabolic modules. To remove this bias and identify specific modules whose enrichment was highly conserved across microbial populations associated with IBD, we implemented two criteria: 1) we selected modules that passed a high statistical significance threshold in our enrichment test (Wilcoxon Rank Sum Test, FDR-adjusted p-value < 2e-10), and 2) we accounted for effect size by ranking these modules according to the difference between their median PPCN in IBD samples and their median PPCN in healthy samples, and keeping only those in the top 50% (which translated to an effect size threshold of > 0.12).

This analysis revealed a set of metabolic modules that were consistently and highly significantly enriched in microbial communities associated with IBD. The majority of these metabolic modules encode biosynthesis pathways. Our use of the terms “elevated”, “enriched”, and “significantly enriched” in the previous version of the text was confusing to the reader. We thank the reviewer for pointing this out, and we hope that our revision of the text clarifies the analysis strategy and observations:

“To gain insight into potential metabolic determinants of microbial survival in the IBD gut environment, we assessed the distribution of metabolic modules within samples from each group (IBD and healthy) with and without using PPCN normalization. Without normalizing, module copy numbers were overall higher in healthy samples (Figure 2a) and modules exhibited weak differential occurrence between cohorts (Figure 2b, 2c, Supplementary Figure 3). The application of PPCN reversed this trend, and most metabolic modules were elevated in IBD (Supplementary Figure 5). This observation is influenced by two independent aspects of the healthy and IBD microbiota. The first one is the increased representation of microbial organisms with smaller genomes in healthy individuals (Watson et al. 2023), which increases the likelihood that the overall copy number of a given metabolic module is below the actual

number of populations. In contrast, one of the hallmarks of the IBD microbiota is the generally increased representation of organisms with larger genomes (Watson et al. 2023). The second aspect is that the generally higher diversity of microbes in healthy individuals increases the denominator of the PPCN. This results in a greater reduction in the PPCN of metabolic modules that are not shared across all members of the diverse gut microbial populations in health.

To go beyond this general trend and identify modules that were highly conserved in the IBD group, we first selected those that passed a relatively high statistical significance threshold in our enrichment test (Wilcoxon Rank Sum Test, FDR-adjusted p-value < 2e-10). We then accounted for effect size by ranking these modules according to the difference between their median PPCN in IBD samples and their median PPCN in healthy samples, and keeping only those in the top 50% (which translated to an effect size threshold of > 0.12). This stringent filtering revealed a set of 33 metabolic modules that were significantly enriched in metagenomes obtained from individuals diagnosed with IBD (Figure 2d, 2e), 17 of which matched the modules that were associated with high metabolic independence previously (Watson et al. 2023) (Figure 2f). This result suggests that the PPCN normalization is an important step in comparative analyses of metabolisms between samples with different levels of microbial diversity."

Lines 178-185 from our original submission have been removed to avoid further confusion. These results can be found in Supplementary File 1 (section "Module enrichment without consideration of effect size leads to nonspecific results").

> *It is not entirely clear to me what is meant by PPCN normalization. Normalize the number of copy numbers to the overall number of genes?*

The idea behind using per-population copy number (PPCN) is to normalize the prevalence of each metabolic module found in an environment with the number of microbial populations within the same sample. PPCN achieves this by dividing the pathway copy numbers by the number of microbial populations in a given metagenome, which we estimate from the frequency of bacterial single-copy core genes. We have updated the description of the per-population copy number (PPCN) calculation to clarify its use:

"Briefly, the PPCN estimates the proportion of microbes in a community with a particular metabolic capacity (Figure 1, Supplementary Figure 2) by normalizing observed metabolic module copy numbers with the 'number of microbial populations in a given metagenome', which we estimate using the single-copy core genes (SCGs) without relying on the reconstruction of individual genomes."

We also note that the equation for PPCN is shown in Figure 1.

> *It is also not clear to me how the classifier predicts stress on microbiomes rather than dysbiosis.*

The reviewer asks an interesting question since it is true that we could also use the term "dysbiosis" rather than "stress". Yet we refrained from the use of dysbiosis as it is considered a poorly-defined term to describe an altered microbiome often associated with a specific disease (doi:https://doi.org/10.3390/microorganisms10030578), such as IBD, relative to another poorly-defined state, "healthy microbiome" (doi:https://doi.org/10.1002/phar.2731). We do consider that stress is not necessarily a term that is less vague than dysbiosis, yet it has the advantage of being more common in studies of ecology compared to dysbiosis. Our relatively neutral stance towards which term to use has shifted dramatically due to one critical observation in our study: the identical patterns of enrichment of HMI microbes in individuals diagnosed with IBD as well as in healthy individuals treated with antibiotics. We appreciate

that the observed changes in the antibiotics case can also fulfill the definition of “dysbiosis”, but the term “stress response” more accurately describes what the classifier identifies in our opinion.

> *What is the advantage of using the estimate-metabolism pipeline presented in this article over workflows such as those using genome-scale models, which are repeatedly cited and discussed?*

Genome-scale models are often appropriate for a big-picture view of metabolism, and especially when the capability to perform quantitative simulations like flux-balance analysis is needed. For our investigation, we wanted a more specific and descriptive summary of metabolic capacity, so we focused on individual KEGG modules, which qualitatively describe subsets of the vast metabolic network with pathway names that all readers can understand, rather than working with an abstract model of the entire network. Furthermore, genome-scale models would have prevented us from assessing the redundancy (copy number) of metabolic pathways, as these networks usually focus on the presence-absence of gene annotations for enzymes in the network rather than the copy number of these annotations. The copy number metric has been critical for our analyses, considering that we are focusing on metabolic capacity at the community level and require the ability to normalize this metabolic capacity by the size of the community described by each metagenome. Finally, assessing a discrete set of metabolic pathways yielded a corresponding set of features that we used to create the machine learning classifier, whereas data from genome-scale models would not be as easily transferable into classifier features.

> *Minor comments:*
>
> *Figure 2d and e are mentioned in the text before Figure 2a.*

We thank the reviewer for catching this. We have rewritten the section as follows to put the figure references in numerical order:

!To gain insight into potential metabolic determinants of microbial survival in the IBD gut environment, we assessed the distribution of metabolic modules within samples from each group (IBD and healthy) with and without using PPCN normalization. Without normalizing, module copy numbers were overall higher in healthy samples (Figure 2a) and modules exhibited weak differential occurrence between cohorts (Figure 2b, 2c, Supplementary Figure 3). After the application of PPCN, most metabolic modules were elevated in IBD (Supplementary Figure 5). This observation is a product of two independent aspects of the healthy and IBD microbiota. The first one is the increased representation of microbial organisms with smaller genomes in healthy individuals (Watson et al. 2023), which increases the likelihood that the overall copy number of a given metabolic module is below the actual number of populations. In contrast, one of the hallmarks of the IBD microbiota is the generally increased representation of organisms with larger genomes (Watson et al. 2023). The second aspect is that the generally higher diversity of microbes in healthy individuals increases the denominator of the PPCN due to the higher number of populations detected in these samples. This results in a greater reduction in the PPCN of metabolic modules that are not shared across all members of the diverse gut microbial populations in health. To go beyond this general trend and identify modules that were highly conserved in the IBD group, we first selected those that passed a relatively high statistical significance threshold in our enrichment test (Wilcoxon Rank Sum Test, FDR-adjusted p-value <2e-10). We then accounted for effect size by ranking these modules according to the difference between their median PPCN in IBD samples and their median PPCN in healthy samples, and keeping only those in the top 50% (which translated to an effect size threshold of > 0.12). This stringent filtering revealed a set of 33 metabolic modules that were significantly enriched in metagenomes obtained from individuals diagnosed with IBD (Figure 2d, 2e), 17 of which matched the

modules that were associated with high metabolic independence previously (Watson et al. 2023) (Figure 2f). This result suggests that the PPCN normalization is an important step in comparative analyses of metabolisms between samples with different levels of microbial diversity.!

> *How much preparation is needed for users that want to apply the estimate-metabolism pipeline to their own datasets? From the documentation at anvi'o, it still seems like a significant effort.*

We thank the reviewer for this important question. The use of anvi-estimate-metabolism is simple, but the concept it makes available and the means it offers its users to interact with their data are not basic, thus its use requires *some* effort. Anvi'o provides users with the ability to directly interact with their data at each step of the analysis to have full control over the analysis and to make informed decisions on the way. In comparison to pre-defined analysis pipelines that often require no additional input from the user, this approach requires some level of involvement of the user throughout the process – namely, they must run a few programs in series rather than running just one pipeline command that quietly handles everything on their behalf. The most basic workflow for using ànvi-estimate-metabolism ìs quite straightforward and requires four simple steps following the installation of anvi'o: 1. Run the program ànvi-setup-kegg-data ùto download the KEGG data. 2. Convert the assembly FASTA file into an anvi'o-compatible database format with gene calls by running ànvi-gen-contigs-database.`3. Annotate genes with KOs with the program ànvi-run-kegg-kofams.`4. Get module completeness scores and copy numbers by running ànvi-estimate-metabolism.`In addition, we provide simple tutorials (such as the one at https://anvio.org/tutorials/fmt-mag-metabolism/) and reproducible bioinformatics workflows online (including for this study at https://merenlab.org/data/ibd-gut-metabolism/) which helps early career researchers to apply similar strategies to their own datasets. We are happy to report that we have been using this tool in our undergraduate education, and observed that students with no background in computation were able to apply it to their questions without any trouble.

> ***Reviewer 2:***
>
> *Congratulations on this great work, the manuscript is a pleasure to read. Minor questions that the authors might want to clarify:*
>
> *L 275: Why use reference genomes from the GTDB (for only 3 phyla) instead of using MAGs reconstructed from the data? I understand that assemblies based on individual samples would probably not yield enough complete MAGs, but I would expect that co-binning the assemblies for the entire dataset would.*

We thank the reviewer for their kind words. We certainly agree that metagenome assembled genomes (MAGs) reconstructed directly from the assemblies would by nature represent the populations in these communities better than reference genomes. However, one of our aims in this study was to avoid the often error-prone and time-consuming step of reconstructing MAGs. Most automatic binning algorithms inevitably make mistakes, and especially for metabolism estimation, low quality MAGs can introduce a bias in the analysis. At the same time the manual curation of each bin to remove any contamination would require a substantial effort and make the workflow less accessible for others to use. As an example, in our previous work (doi:10.1186/s13059-023-02924-x), careful refinement of MAGs from just two co-assemblies took two months. Here, we developed the PPCN workflow as a more scalable, assembly-level analysis to avoid the need for binning in the first place.

To supplement and confirm the metagenome-level results, we decided to run a genome-level analysis. We used the GTDB since it represents the most comprehensive, dereplicated

collection of reference genomes across the tree of life. We chose those 3 phyla in particular because of their ecological relevance in the human gut environment. Bacteroidetes and

Firmicutes together represent the majority (up to ~90%) of the populations in healthy individuals (doi:10.1038/nature07540), and Proteobacteria represent the next most abundant phylum on average (2% ± 10%) (doi:10.1371/journal.pone.0206484).

> *L 403: Should the Franzosa and Papa papers be referenced as numbers?*

Thanks for pointing this out. The rogue numerical citation was actually an artifact of the submission and was corrected to a long-format citation in the online version of the manuscript on the eLife website.

> ***Reviewer 3:***
>
> *The lack of any experimental validation contributes to the tentative nature of the conclusions that can be drawn at this time. Numerous studies have looked at the metabolism of gut bacterial species during in vitro growth, which could be mined to test if the in silico predictions of metabolism can be supported. Alternatively, the authors could isolate key strains of interest and study them in culture or in mouse models of IBD.*

We appreciate these suggestions and agree with the reviewer that experimental validation is important. However, we do not agree that either the use of mouse models or the isolation of individual microbial strains would be an appropriate experimental test in this case. The use of humanized gnotobiotic mice has critical limitations (see doi:10.1016/j.cell.2019.12.025 and references within the section on “human microbiota-associated murine models”). As it is not possible to establish a mouse model whose gut microbiota fully reflect the human gut microbiome, such an approach would neither be appropriate to validate our findings, nor would it have been possible to produce the insights we have gained based on environmental data. We are not sure how exactly a mouse model, even when ignoring the well established limitations, could improve or validate a comprehensive analysis of a large “environmental” datasets that resulted in highly significant signals.

We are also not sure that we understand how the reviewer believes that the isolation of individual strains would aid in validating our findings. While we appreciate that not all relevant genes are captured by the available annotation routines and that some genes may be misannotated, the large dataset used here renders these concerns negligible. Isolating a small subset of bacterial populations would hardly lead to a representative sample and testing their metabolic capacities in vitro would not improve the reliability of our analysis.

Boilerplate suggestions as vague as “isolate key strains of interest” or “experiment in mouse models of IBD” do not add or retract anything from our findings. Our findings and hypotheses are well supported by our data and extensive analyses.

> *Line 9 - not sure this approach is hypothesis testing in the traditional sense, you might reword.*

Hypothesis testing occurs when one makes an observation, develops an hypothesis that explains the observation, and then gathers and analyzes data to investigate whether additional data support or disprove the hypothesis. We are not convinced a reword is necessary.

> *Line 40 - the lack of consistent differences in IBD and healthy individuals does not mean that the microbiome doesn't impact disease. It's important to consider all the mechanistic studies in animal models and other systems.*

Our study does not claim that microbiome has no impact on the course of disease.

> *Line 50 - this seemed out of place and undercuts the current findings. Upon checking Ref. 31, the analysis seems distinct enough to not mention in the introduction.*

We disagree. Ref 31 uses genome-scale metabolic models to identify the loss of cross-feeding interactions in the gut microbiome of individuals with IBD, which is another way of saying that the microbes in IBD no longer rely on their community for metabolic exchange – in other words, they are metabolically independent. This is an independent observation that is parallel to our results and confirms our analysis; hence, it is important to keep in our introduction.

> *Line 55 - Ref. 32 looked at FMT, which should be explicitly stated here.*

The reviewer's suggestion is not helpful. Ref 32 has a significant focus on IBD as it compares a total of 300 MAGs generated from individuals with IBD to 264 MAGs from healthy individuals and shows differences in metabolic enrichment between healthy and IBD samples independent of taxonomy, thus setting the stage for our current work. What model has been used to generate the initial insights that led to the IBD-related conclusion in Ref 32 has no significance in this context.

> *Lines 92-107 - this text is out of place in the Results section and reads more like a review article. Please trim it down and move it to the introduction.*

We would like to draw the reviewer's attention to the fact that this is a "Result and Discussion" section. In this specific case it is important for readers to appreciate the context for our new tool, as the reviewer commented in the public review. We kindly disagree with the reviewer's suggestion to remove this text as that would diminish the context.

> *Line 107 - is "selection" the word you meant to use?*

If the frequency of a given metabolic module remains the same or increases despite the decreasing diversity of the microbial community, it is conceivable to assume that its enrichment indicates the presence of a selective process to which the module responds. It is indeed the word we meant to use.

Line 110 - this is the first mention of this new method, need to add it to the abstract and introduction.

The reviewer must have overlooked the text passages in which we mention the strategy we developed within the abstract:

"Here, we tested this hypothesis on a large scale, by developing a software framework to quantify the enrichment of microbial metabolisms in complex metagenomes as a function of microbial diversity."

And in the last paragraph of the introduction:

"Here we implemented a high-throughput, taxonomy-independent strategy to estimate metabolic capabilities of microbial communities directly from metagenomes…"

> *Figure 1 - a nice summary, but no data is shown to support the validity of this model. Consider shrinking the cartoon and adding validation with simulated datasets.*

We hope we have addressed this recommendation with the extensive validation efforts summarized above.

> *Line 134 - need to state the FDR and effect size cutoffs used.*

We have reworded this sentence as follows to clarify which thresholds were used:

“We identified significantly enriched modules using an FDR-adjusted p-value threshold of p < 2e-10 and an effect size threshold of > 0.12 from a Wilcoxon Rank Sum Test comparing IBD and healthy samples.”

> *I'm also concerned about the simple comparison of IBD to healthy without adjusting for confounders like study, geographical location, age, sex, drug use, diet, etc. More text is needed to explain the nature of these data, how much metadata is available, and which other variables distinguish IBD from healthy.*

The reviewer is correct that there is a large amount of interindividual variation between samples due to host and environmental factors. However, the lack of adjusting for confounders was intentional, and in fact one of the critical strengths of our study. We observe a clear signal between healthy individuals and individuals diagnosed with IBD, *despite* the amount of interindividual variation in our diverse set of samples from 13 different studies (details of which are summarized in Supplementary Table 1). The clear increase in predicted metabolic capacity that we consistently observe in IBD patients using both metagenomes and genomes across diverse cohorts points to metabolic independence as a high-level trend that is predictive of microbial prevalence in stressed gut environments irrespective of host factors.

> *Line 145 - calling PPCN normalization an "essential step" is a huge claim and requires a lot more data to back it up. Might be best to qualify this statement.*

We hope we have addressed this recommendation with our validation efforts. Supplementary Figures 18 and 19 in particular show evidence for the necessity of the normalization step. It is indeed an essential step *if* the purpose is to compare metabolic enrichment between cohorts of highly different microbial diversity.

> *Figure 2a - the use of a 1:1 trend line seems potentially misleading. I would replace it with a best-fit line.*

Our purpose here was not to show the best fit. Instead, the 1:1 trend line separates the modules based on their relative abundance distribution between healthy individuals and individuals diagnosed with IBD. If the module is to the left of the line, it has a higher median copy number in healthy individuals and if the module is to the right, it has a higher median copy number in individuals with IBD. The line also helps to demonstrate the shift that occurs between the unnormalized data in Figure 2a. Without the normalization, more modules occur to the left of the

1/1 line as a result of the higher raw copy numbers in healthy metagenomes which simply contain more microbial populations. With the normalization (Figure 2d), more modules fall on the right side of the 1/1 line due to higher PPCN values. A best-fit line would not serve well for these purposes.

> *The text should be revised to state that this analysis actually did find many significant differences and to discuss whether they were the same modules identified in Figure 2d.*

We apologize for the confusion and thank the reviewer for bringing this issue to our attention. As mentioned above, the disparate levels of microbial diversity between healthy individuals and individuals with IBD resulted in much larger copy numbers of metabolic modules in healthy samples reflecting the often much larger communities. Hence, we ran statistical tests only on normalized (PPCN) data. The p-values associated with each module in Figure 2a, as well as the colors of each point, are based on the PPCN data in Figure 2d. We aimed to improve the clarity of the visual comparison between normalized and unnormalized results by identifying the same set of IBD-enriched modules in plots a-c and plots d-f.

That being said, the reviewer's comment made us realize the potential for confusion when using the normalized data's statistical results in Figure 2a that otherwise shows results from unnormalized data. We have now run the same statistical test on the unnormalized (raw copy number) data and re-generated Figure 2a with the new FDR-adjusted p-values and points colored based on the statistical tests using unnormalized data. We've also removed the arrow connecting to Figure 2b (since we no longer show the same set of IBD-enriched modules in Figures 2a and 2b), and added a dashed line to indicate the effect size threshold (similar to the one in Figure 2d). We have updated the legend for Figure 2a-d to reflect these changes:

When we used the same p-value threshold ($p < 2e\text{-}10$) as before and also filtered for an effect size larger than the mean (the same strategy used to set our effect size threshold for the normalized data), there are 10 modules that are significantly enriched based on the unnormalized data. Of course, it is difficult to gauge the relevance of these 10 modules to microbial fitness in the IBD gut environment since their raw copy numbers do not tell us anything about the relative proportion of community members that harbor these modules. Therefore, we are reluctant to add these modules to the results text. For the record, only 3 of those modules were also significantly enriched based on the normalized PPCN values: M00010 (Citrate cycle, first carbon oxidation), M00053 (Pyrimidine deoxyribonucleotide biosynthesis), and M00121 (Heme biosynthesis).

> *Figure 2c,f - these panels raise a lot of concerns given that the choice of method inverts the trend. Without additional data/validation, it's hard to know which method is right.*

We hope we have addressed this recommendation with the extensive validation efforts summarized above. Inversion of the trend is an expected outcome, because the raw copy numbers of most metabolic modules are much lower in the IBD sample group due to lower community sizes.

> *Line 167 - Need to take the KEGG names with a grain of salt, just because it says "biosynthesis" doesn't mean that the pathway goes in that direction in your bacterium of interest.*

We believe the reviewer is under a misapprehension regarding the general reversibility of KEGG metabolic modules, or indeed of metabolic pathways. Most metabolic pathways have one or several (practically) irreversible reactions. To demonstrate this for the 33 IBD-enriched modules, we evaluated their reversibility based upon their corresponding KEGG Pathway Maps, which indicate reaction reversibility via double-sided arrows. Aside from the signature modules M00705 and M00627, in 26 out of 31 pathway modules one or more irreversible reactions render these pathways one-directional. Indeed, on average the majority (54%) of the reactions in a given module are irreversible. When focusing on the 23 "biosynthesis" modules, 22 out of 23 (96%) modules have at least one irreversible reaction, and on average 64% of a given module's reactions are irreversible. These data (which can be accessed at doi:10.6084/m9.figshare.27203226 for the reviewer's convenience) challenge the reviewer's notion that pathway directionality is free to change arbitrarily, since the presence of even one

irreversible reaction effectively blocks the flux in the opposing direction. Thus, “biosynthesis” is indeed a meaningful term in KEGG module names.

That said, KEGG Pathway Maps, though highly curated, are likely not the final word on whether a given reaction in a metabolic pathway can be considered reversible or irreversible in each microbial population and under all conditions. And our analysis, like many others that rely on metagenomic data, does not consider the environmental conditions in the gut such as temperature or metabolite concentrations that might influence the Gibbs free energy and thus the directionality of these reactions in vivo. However, even assuming general reversibility of metabolic pathways, this would not invalidate the fact that these microbes have the metabolic capacity to synthesize the respective molecules. In other words, the potential reversibility of pathways is irrelevant to our analysis since we are describing metabolic *potential*. The *lac* operon in *E. coli* might only be expressed in the absence of glucose, but *E. coli* always has the capability to degrade lactose regardless of whether that pathway is active. Thus, our overall conclusion that gut microbes associated with IBD are metabolically self-sufficient (encoding the enzymatic capability to synthesize certain key metabolites) remains valid irrespective of fixed or flexible pathway directionality.

> *It's also important to be careful not to conflate KEGG modules (small subsets of a pathway) with the actual metabolic pathway. It's possible to have a module change in abundance while not altering the full pathway. Inspection of the individual genes could help in this respect - are they rate-limiting steps for biosynthesis or catabolism?*

The reviewer is absolutely correct that KEGG modules do not necessarily represent full pathways. We have updated the language in our manuscript to explicitly refer to “modules” rather than “pathways” whenever appropriate, to restrict the scope of the analysis to metabolic modules rather than full pathways.

That said, we do not see how “inspection of individual genes” would improve our analysis. The strength of looking at complete modules rather than individual genes is that we can gain conclusive insights into a certain metabolic capacity. Of course, no pathway or module stands alone. However, the enrichment of metabolic modules does conclusively indicate that these modules are beneficial under the given conditions, such as stress caused by inflammation or antibiotic use. Whether a certain step in a module or pathway is rate limiting is completely irrelevant for this analysis.

> *Line 177 - I'm not a big fan of the HMI acronym. Is there a LMI group? It seems simplistic to lump all of metabolism into dependent or independent, which in reality will differ depending on the specific substrate, the growth condition, and the strain.*

While we are sorry that our study failed to provide the reviewer with a term they could be a fan of, their input did not change our view that HMI, an acronym we have adapted from a previously peer-reviewed study (doi:10.1186/s13059-023-02924-x), is a powerfully simplistic means to describe a phenomenon we observe and demonstrate in multiple different ways with our extensive analyses. The argument that HMI or LMI status will differ given the growth condition, substrate availability, or strain differences is not helping this case either: our analyses cut across a large number of humans and naturally occurring microbial systems in their guts that are exposed to largely variable ‘growth conditions’ and ‘substrates’ and composed of many strain variants of similar populations. Yet, we observe a clear role for HMI despite all these differences. Perhaps it is because HMI simply describes a higher metabolic capacity based on a defined subset of largely biosynthetic pathways that we observe to be consistently enriched in a large dataset covering a large variety of host, environmental and diet factors and indicates that a population has a higher metabolic capacity to not rely on ecosystem services. We show in our analysis that in the inflamed gut these capacities are indeed required, which is why HMI populations are enriched in IBD samples. HMI has no

relation to any of the constraints mentioned by the reviewer, which is one of the major strengths of this metric.

> *Line 198 - It seems like a big assumption to state that efflux and drug resistance are unrelated to biosynthesis, as they could be genetically or even phenotypically linked.*

We agree with the reviewer and are thankful for their input. We have weakened the assertion in this statement.

“These capacities may provide an advantage since antibiotics are a common treatment for IBDs (Nitzan et al. 2016), but are not necessarily related to the systematic enrichment of biosynthesis modules that likely provide resilience to general environmental stress rather than to a specific stressor such as antibiotics.”

> *Lines 202-218 - I'd suggest removing this paragraph. The "non-IBD" data introduces even more complications to the meta-analysis and seems irrelevant to the current study.*

We thank the reviewer for this suggestion. Non-IBD data is important, but its relevance to the primary aims of the study is indeed negligible. We now have moved this paragraph to Supplementary File 1 (under the section “‘Non-IBD’ samples are intermediate to IBD and healthy samples”).

> *The health gradient is particularly problematic, putting cancer closer to healthy than IBD.*

We took the reviewer’s advice and have swapped the order of the studies in Supplementary Figure 6 to place the cancer samples from Feng et al. closer to the IBD samples, on the other side of the non-IBD samples from the IBD studies.

> *Lines 235-257 - should trim this down and move to the discussion.*

As mentioned above, we have opted for a “Results and Discussion format” for our manuscript, so we believe this discussion is in the correct place. We find it important to clearly highlight the limitations and potential biases of our work and trimming this text would take away from that goal.

> *Figure 3 - panels are out of order. Need to put the current panel D below current panel C. Also, relabel panel letters to go top to bottom (the bottom panel should be D). Could change current panel 3D to a violin plot to match current 3C.*

We have updated Figure 3 by converting panel A into a new supplementary figure (Supplementary Figure 8), moving panels C and D below panel B, and relabeling the panels accordingly.

> *Figure 3B - this panel was incredibly useful and quite surprising to me in many respects. I would have assumed that the Bacteroides would be in the "HMI" bin. Is this a function of the specific strains included here? Was B. theta or B. fragilis included?*

The reviewer makes an excellent observation that has been keeping us awake at night, yet somehow was not appropriately discussed in the text until their input. We are very thankful for their attention to detail here.

It is indeed true that *Bacteroides* genomes are often detected with increased abundance in individuals with IBD and likely have a survival advantage in the IBD gut environment, *Bacteroides fragilis* and *Bacteroides thetaiotaomicron* being some of the most dominant residents of the IBD gut. Their non-HMI status is not a function of which strains were

included, since all taxa here are represented by the representative genomes available in the publicly available Genome Taxonomy Database. Their non-HMI status comes from the fact that they have HMI scores of around 24 to 26, which fall slightly below the threshold score of 26.4 that we used to classify genomes as HMI. This threshold is back-calculated from the metabolic completion requirement of at least 80% average completion of all 33 metabolic modules that are significantly enriched in IBD. So these genomes are right there at the edge, but not quite over it.

Thanks to this comment by our reviewer, we started wondering whether we should follow a more 'literature-driven' approach to set the threshold for HMI, rather than the 80% cutoff, and in fact attempted to lower the HMI score threshold to see if we could include more of the IBD-associated *Bacteroides* in the HMI bin. Author response table 1 below shows the relevant subset of our new Supplementary Table 3h, which describes the data from our tests on different thresholds.

Author response table 1.

Number and proportion of *Bacteroides* genomes classified as HMI at each HMI score threshold. There were 20 total *Bacteroides* genomes in the set of 338 gut microbes identified from the GTDB. The HMI score is computed by adding the percent completeness of all 33 IBD-enriched KEGG modules. The full table can be viewed in Supplementary Table 3h.

Average percent completeness of IBD-enriched modules	Corresponding HMI score threshold	Number of *Bacteroides* genomes classified as HMI	Percent of *Bacteroides* genomes classified as HMI
75	24.75	6	30%
76	25.08	4	20%
77	25.41	3	15%
78	25.74	0	0%
79	26.07	0	0%
80	26.4	0	0%

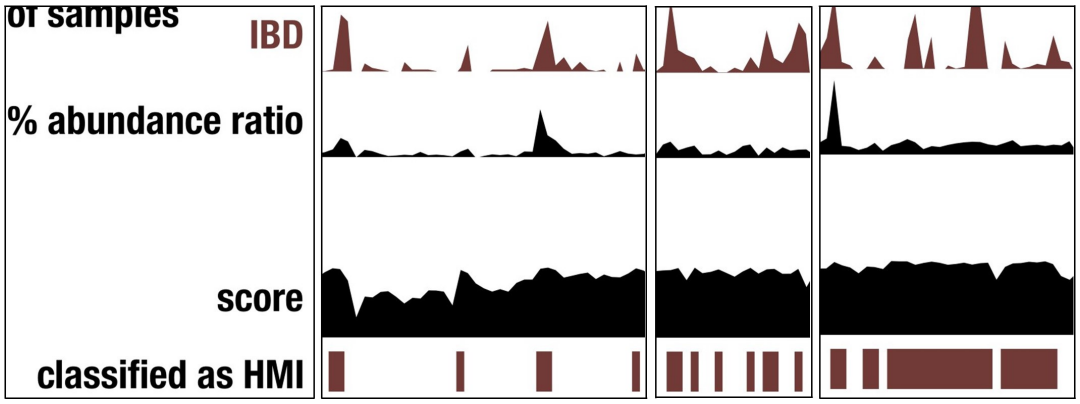
Lowering the threshold to 24.75, which corresponds to an average of 75% completeness in the 33 IBD-enriched modules, enabled the classification of 6 *Bacteroides* genomes as HMI, including *B. fragilis*, *B. intestinalis*, *B. theta,* and *B. faecis*. However, it also identified several microbes that are not IBD-associated as HMI, including 75 genomes from the Lachnospiraceae family and 18 genomes from the Ruminococcaceae family. In the latter family, several *Faecalibacterium* genomes, including 10 representatives of *Faecalibacterium*

prausnitzii, were considered HMI using this threshold. These microbes are empirically known to decrease in abundance during inflammatory gastrointestinal conditions (doi:10.3390/microorganisms8040573, doi:10.1093/femsre/fuad039), and therefore these genomes should not be considered HMI – at least not under the working definition of HMI used in our study. To avoid including such a large number of obvious false positives in the HMI bin, we decided to maintain a higher threshold despite the exclusion of *Bacteroides* genomes.

This outcome demonstrates that our reductionist approach does not successfully capture every microbial population that is associated with IBD. Nevertheless, and in our opinion very surprisingly, the metric does capture a very large proportion of genomes with increased detection and abundance in IBD samples, as demonstrated by the peaks of detection/abundance that match to HMI status Author response image 1.

Author response image 1.

Screenshots of Figure 3 that demonstrate the overlapping signal between HMI status and genome detection/abundance in IBD.



Furthermore, the violin plots in Figure 3B (formerly Figure 3C) clearly reflect the increased representation of HMI populations in IBD metagenomes. Although our classification method is imperfect, it still demonstrates the predictive power of metabolic competencies in identifying which microbes will survive in stressful gut environments. To ensure that readers recognize the crude nature of this classification strategy and the possibility that high metabolic independence can be achieved in different ways, we have added the following sentences to the relevant section of our manuscript:

“Given the number of ways a genome can pass or fail this threshold, this arbitrary cut-off has significant shortcomings, which was demonstrated by the fact that several species in the Bacteroides group were not classified as HMI despite their frequent dominance of the gut microbiome of individuals with IBD (Saitoh et al. 2002; Wexler 2007; Vineis et al. 2016) (Supplementary File 1). That said, the genomes that were classified as HMI by this approach were consistently higher in their detection and abundance in IBD samples (Figure 3a). It is likely that there are multiple ways to have high metabolic independence which are not fully captured by the 33 IBD-enriched metabolic modules identified in this study.”

We have also included a discussion of these findings in Supplementary Information File 1 (see section “Examining the impact of different HMI score thresholds on genome-level results”).

> *This panel also makes it clear that many of these modules are widespread in all genomes and thus unlikely to meaningfully differ in the microbiome. It would be interesting to use*

> *this type of analysis to identify a subset of KEGG modules with high variability between strains.*

The figure makes it 'look like' many of these modules are widespread in all genomes and thus unlikely to meaningfully differ in the microbiome, but our quantitative analyses clearly demonstrate that these modules indeed differ meaningfully between microbiomes of healthy individuals and those diagnosed with IBD. For instance, the classifier that we built relying exclusively upon these modules' PPCN values was able to reliably distinguish between the healthy and IBD sample groups in our dataset. The fact that the differentiating signal does not rely on rare metabolic or signature modules is what makes the classifier powerful enough to differentiate between "healthy" and "stressed" microbiomes in 86% of cases. Modules that are by nature less common could not serve this purpose. That said, we do agree with the reviewer that it might be interesting to study variability of KEGG modules as a function of variability between strains. This does not fall into the scope of this work, but we hope to assist others with the technical aspects of such work.

Considering the entirety of the exchange in this section, perhaps there is a broader discussion to be had around this topic. In retrospect, not being able to perfectly split microbes into two groups that completely recapitulate their enrichment in healthy or IBD samples by a crude metric and an arbitrary threshold is not surprising at all. What is surprising is that such a crude metric in fact works for the vast majority of microbes and predicts their increased presence in the IBD gut by only considering their genetic make up. In some respects, we believe that the inability of this cutoff to propose a perfect classifier is similar to the limited power of metabolic independence concept and the classes of HMI or LMI to capture and fully explain microbial fitness in health and disease. What is again surprising here is that these almost offensively simple classes do capture more than what one would expect. We can envision a few ways to implement a more sophisticated HMI/LMI classifier, and it is certainly an important task that is achievable. However, we are hopeful that this technical work can also be done better by others in our field, and that step forward, along with further scrutinizing the relevance of HMI/LMI classes to understand metabolic factors that contribute to the biodiversity of stressful environments, will have to remain as future work.

We thank the reviewer again for their comment here and pushing us to think more carefully and address the oddity regarding the poor representation of *Bacteroides* as HMI by our cutoff.

> *Given that a lot of the gaps are in the Firmicutes, this panel also makes me more concerned about annotation bias. How many of these gaps are real?*

Analyses relying on gene annotations all suffer equally from the potential for missannotation or missing annotations, which primarily result from limitations in our reference databases for functional data. For instance, the Hidden Markov models for microbial genes in the KEGG Ortholog database are generated from a curated set of gene sequences primarily originating from cultivable microorganisms and particularly from commonly-used model organisms; hence, they do not capture the full extent of sequence diversity observed in populations that are less well-represented in reference databases – a category which includes several Firmicutes, as the reviewer points out. For KEGG KOfams in particular, the precomputed bit score thresholds for distinguishing between 'good' and 'bad' matches to a given model are often too stringent to enable annotation of genes that are just slightly too divergent from the set of known sequences, thus resulting in missing annotations. Based on our experience with these sorts of issues, we implemented a heuristic that reduces the number of missing annotations for KOs and captures significantly more homologs than other state-of-the-art approaches, as described in doi:10.1101/2024.07.03.601779. We refer the reviewer to our response to the related public comment about annotation bias above, which includes additional details about our investigations of annotation bias in our data. In comparison to the current standard, the heuristic we implemented improves functional annotation results.

However, neither our nor any other bioinformatic study that relies on functional gene annotation can exclude the potential for annotation bias.

> *Figure 3B plotting issues - need to use the full names of the modules; for example, M00844 is "arginine biosynthesis, ornithine => arginine", which changes the interpretation. Need a key for the heatmap on the figure. The tree is difficult to see, needs a darker font.*

We have darkened the lines of the tree and dendrogram, and added a legend for the heatmap gradient (see new version of Figure 3 above). Unfortunately, we could not fit the full names of the modules into the figure due to space constraints. However, the full module name and other relevant information can be found in Supplementary Table 2a, and the matrix of pathway completeness scores in these genomes (e.g., the values plotted in the heatmap) can be found in Supplementary Table 3b. We are not sure what the reviewer refers to when stating that “for example, M00844 is "arginine biosynthesis, ornithine => arginine", which changes the interpretation”. There is no ambiguity regarding the identity of KEGG module M00844, which is arginine biosynthesis from ornithine.

> *Line 321 - more justification for the 80% cutoff is needed along with a sensitivity analysis to see if this choice matters for the key results.*

Inspired by this comment, and the one above regarding the classification of *Bacteroides* genomes, we tested several HMI score thresholds ranging from 75% to 85% average completeness of the 33 IBD-enriched modules. For each threshold, we computed all the key statistics reported in this section of our paper, including the statistical tests. We found that the choice of HMI score threshold does not influence the overall conclusions drawn in this section of our manuscript. Author response table 2 below shows the relevant subset of our new Supplementary Table 3h, which describes the results for each threshold:

Author response table 2.

Key genome-level results at each HMI score threshold. The HMI score is computed by adding the percent completeness of all 33 IBD-enriched KEGG modules. WRS – Wilcoxon Rank Sum test; KW – Kruskal-Wallis test. The full table can be viewed in Supplementary Table 3h

Average percent completeness of IBD-enriched modules	HMI score threshold	Number of HMI genomes	WRS p-value for HMI vs non-HMI detection in IBD	WRS p-value for HMI vs non-HMI detection in healthy	KW p-value for fraction of HMI in IBD vs non-IBD samples
75	24.75	129	2.18E-08	0.001776512	1.24E-16
76	25.08	115	2.70E-08	0.002832713	2.54E-19
77	25.41	98	2.10E-06	0.012257189	1.60E-18
78	25.74	78	4.16E-06	0.087182579	4.36E-21
79	26.07	69	5.46E-06	0.078132573	1.75E-21
80	26.4	59	4.82E-06	0.267265165	8.98E-25
81	26.73	48	0.000168971	0.700660484	5.03E-25
82	27.06	39	0.000217924	0.836279996	2.89E-29
83	27.39	35	0.001750966	0.962660229	3.77E-30
84	27.72	28	0.024916547	0.997986302	5.66E-30
85	28.05	24	0.043823626	0.999220691	5.21E-30

We've summarized these findings in a new section of Supplementary File 1 entitled "Examining the impact of different HMI score thresholds on genome-level results". We copy below the relevant text for the reviewer's convenience:

"Determining the HMI status of a given genome required us to set a threshold for the HMI score above which a genome would be considered to have high metabolic independence. We tested several different thresholds by varying the average percent completeness of the 33 IBD-enriched metabolic modules that we expected from the

'HMI' genomes from ≥ 75% (corresponding to an HMI score of ≥ 24.75) to ≥ 85% (corresponding to an HMI score of ≥ 28.05). For each threshold, we computed the same

statistics and ran the same statistical tests as those reported in our main manuscript to assess the impact of these thresholds on the results (Supplementary Table 3h). At the highest threshold we tested (HMI score ≥ 28.05), a small proportion of the reference genomes (7%, or n = 24) were classified as HMI, so we did not test higher thresholds.

We found that the results from comparing HMI genomes to non-HMI genomes are similar regardless of which HMI score threshold is used to classify genomes into either group. No matter which HMI score threshold was used, the mean genome size and mean number of genes were higher for HMI genomes than for non-HMI genomes. On average, the HMI genomes were about 1 Mb larger and had 1,032 more gene calls than non-HMI genomes. We ran two Wilcoxon Rank Sum statistical tests to assess the following null hypotheses: (1) HMI genomes do not have higher detection in IBD samples than non-HMI genomes, and (2) HMI genomes do not have higher detection in healthy samples than non-HMI genomes. For both tests, the p-values decreased (grew more significant) as the HMI score threshold decreased due to the inclusion of more genomes in the HMI bin. The first test for higher detection of HMI genomes than non-HMI genomes in IBD samples yielded p-values less than $\alpha = 0.05$ at all HMI score thresholds. The second test for higher detection of HMI genomes than non-HMI genomes in healthy samples yielded p-values less than $\alpha = 0.05$ for the three lowest HMI score thresholds (HMI score ≥ 24.75, ≥ 25.08, or ≥ 25.41). However, irrespective of significance threshold and HMI score threshold, there was always far stronger evidence to reject the first null hypothesis than the second, given that the p-value for the first test in IBD samples was 1 to 5 orders of magnitude lower (more significant) than the p-value for the second test in healthy samples.

IBD samples harbored a significantly higher fraction of genomes classified as HMI than healthy or non-IBD samples, regardless of HMI score threshold ($p < 1e\text{-}15$, Kruskal-Wallis Rank Sum test). The p-values for this test increased (grew less significant) as the HMI score threshold decreased. This suggests that, at higher thresholds, relatively more genomes drop out of the HMI fraction in healthy/non-IBD samples than in IBD samples, thereby leading to larger differences and more significant p-values. Consequently, the HMI scores of genomes detected in IBD samples must be higher than the HMI scores of genomes detected in the other sample groups – indeed, the average HMI score of genomes detected within at least one IBD sample is 24.75, while the average score of genomes detected within at least one healthy sample is 22.78. Within a given sample, the mean HMI score of genomes detected within that sample is higher for the IBD group than in the healthy group: the average per-sample mean HMI score is 25.14 across IBD samples compared to the average of 23.00 across healthy samples.”

> *Lines 357 and 454 - I would remove the discussion of the "gut environment" which isn't really addressed here. The observed trends could just as easily relate to microbial interactions or the effects of diet and pharmaceuticals. Perhaps the issue is the vague nature of this term, which I read to imply changes in the mammalian host. Given the level of evidence, I'd opt to keep the options open and discuss what additional data would help resolve these questions.*

We are in complete agreement with the reviewer that microbial interactions are likely an important driver of our observations. In healthy communities, microbial cross-feeding enables microbes with lower metabolic independence to establish and increase microbial diversity. Which is exactly why we are stating that “Community-level signal translates to individual microbial populations and provides insights into the microbial ecology of stressed gut environments”.

Diet or usage of prescription drugs on the other hand, as discussed previously, likely varies substantially over the various cohorts investigated, and is thus not a driver of the observed

trends. Instead, HMI works as a high level indicator that is not influenced by these variable host habits.

> *Lines 354-394 - Could remove or dramatically trim down this text. Too much discussion for a results section.*

We kindly remind the reviewer that our manuscript is written following a “Results and Discussion” format. This section provides necessary context and justification for our classifier implementation, so we have left it as-is.

> *Lines 395-441 - This section raised a lot of issues and could be qualified or even removed. The model was trained on modules that were IBD-associated in the same dataset, so it's not surprising that it worked. An independent test set would be required to see if this model has any broader utility.*

The point that we selected the IBD-enriched modules as features should not raise any concerns, as these modules would have emerged as the most important (ie, most highly weighted) features in our model even if we had included all modules in our training data. This is because machine learning classifiers by design pick out the features that best distinguish between classes, and the 33 IBD-associated modules are a selective subset of these (if they were not, they would not have been significantly enriched in the IBD sample group). That said, a carefully conducted feature selection process prior to model training is a standard best-practice in machine learning; thus, if anything, this should be interpreted as a point of confidence rather than a concern. Furthermore, we evaluated our model using cross-validation, a standard practice in the machine learning field that assesses the stability of model performance by training and testing the model on different subsets of the data. This effort established that the model is robust across different inputs as demonstrated by the per-fold confusion matrix and the ROC curve. These are all standard approaches in machine learning to quantify the model tradeoff between bias and variance. As for the independent test set, we went far and beyond, and applied our model to the antibiotic time-series dataset described later in this section, which, in our opinion, and likely also in the opinion of many experts, serves as one of the most convincing ways to test the utility of any model. Classification results here show that our hypothesis concerning the relevance of metabolic independence to microbial survival in stressed gut environments applies beyond the IBD case and includes antibiotic use, which is indeed a stronger validation for this hypothesis than any test we could have done on other IBD-related datasets. Regardless, we agree that any ‘broader’ utility of our model, such as its applications in clinical settings for diagnostic purposes, is something we certainly can not make strong claims about without more data. We have therefore qualified this section by adding the following sentence:

“Determining whether such a model has broader utility as a diagnostic tool requires further research and validation; however, these results demonstrate the potential of HMI as an accessible diagnostic marker of IBD.”

> *The application to the antibiotic intervention data raises additional concerns, as the model will predict IBD (labeled "stress" in Figure 5) where none exists.*

We apologize for this misunderstanding. The label “stress” actually means stress, not IBD. The figure the reviewer is referring to demonstrates that metabolic modules enriched in the gut microbiome of IBD patients are also temporarily enriched in the gut microbiome of healthy individuals treated with antibiotics for the duration of the treatment. While the classifier uses PPCN values for 33 metabolic modules enriched in microbiomes of IBD patients, it does not mean that this enrichment is exclusive to IBD. The classifier will distinguish between metagenomes in which the PPCN values for those 33 metabolic modules is higher and

metagenomes in which the PPCN values are lower. Hence, our analysis demonstrates that during antibiotic usage in healthy individuals, the PPCN values of these 33 metabolic modules spike *in a similar fashion* to how they would in the gut community of a person with IBD. This points to a more general trend of high metabolic independence as a factor supporting microbial survival in conditions of stress; that is, the increase in metabolic independence is not specific to the IBD condition but rather a more generic ecological response to perturbations in the gut microbial community. We have clarified this point with the following addition to the paragraph summarizing these results:

"All pre-treatment samples were classified as 'healthy' followed by a decline in the proportion of 'healthy' samples to a minimum 8 days post-treatment, and a gradual increase until 180 days post treatment, when over 90% of samples were classified as 'healthy' (Figure 5, Supplementary Table 4b). In other words, the increase in the HMI metric serves as an indicator of stress in the gut microbiome, regardless of whether that stress arises from the IBD condition or the application of antibiotics. These observations support the role of HMI as an ecological driver of microbial resilience during gut stress caused by a variety of environmental perturbations and demonstrate its diagnostic power in reflecting gut microbiome state."

We've also added the following sentence to the end of the legend for Figure 5:

"Samples classified as 'healthy' by the model were considered to have 'no stress' (blue), while samples classified as 'IBD' were considered to be under 'stress' (red)."

> *Figure S5A - should probably split this into 2 graphs since different data is analyzed.*

It is true that different sets of modules are used in either half of the figure; however, there is a significant amount of overlap between the sets (17 modules), which is why there are lines connecting the points for the same module as described in the figure legend. We are using this figure to make the point that the median PPCN value of each module increases, in both sets of modules, from the healthy sample group to the IBD sample group. Therefore, we believe the current presentation is appropriate.

> *Figure S6A – this shows a substantial study effect and raises concerns about reproducibility.*

We examined potential batch effects in Supplementary Information File 1 (see section "Considerations of Batch Effect"), and found that any study effect was minor and overcome by the signal between groups:

"The similar distribution of the median normalized copy number for each of the 33 IBD-enriched metabolic modules (summarized across all samples within a given study), across all studies within a given sample group (Supplementary Figure 6b), confirms that the sample group explains more of the trend than the study of origin."

Furthermore, within Supplementary Figure 6a, there is a clear increase between the non-IBD controls from Franzosa et al. 2018 and the IBD samples from the same study, as well as between the non-IBD controls from Schirmir et al. 2018 and the IBD samples from that study. As there is no study effect influencing those two comparisons, this reinforces the evidence that there is a true increase in the normalized copy numbers of these modules when comparing samples from more healthy individuals to those from less healthy individuals.

> *Figure S7B - check numbers, which I think should sum to 33.*

The numbers should not sum to 33. In this test to determine whether the two largest studies had excessive influence on the identity of the IBD-enriched modules, we repeated our

strategy to obtain 33 IBD-enriched modules (those with the 33 smallest p-values from the statistical test) from each set of samples – either (1) samples from Le Chatelier et al. 2013 and Vineis et al. 2016, or (2) samples that are not from those two studies. The 2 sets, containing 33 modules each, gives us a total of 66 IBD-enriched modules. By comparing those two sets, we found that 20 modules were present in both sets – hence the value of 20 in the center of the Venn Diagram. In each set, 13 modules were unique – hence the value of 13 on either side. 13 + 13 + 2*20 = 66 total modules.

We again thank our reviewers for their time and interest, and invaluable input.

https://doi.org/10.7554/eLife.89862.2.sa0