

Healthy microbiome—moving towards functional interpretation

Kinga Zielińska¹, Klas I. Udekwu^{2,3}, Witold Rudnicki^{4,5}, Alina Frolova^{6,7}, and Paweł P. Łabaj^{1,*}

¹Małopolska Centre of Biotechnology, Jagiellonian University, 30-387 Krakow, Poland

²Department of Biological Sciences, Bioinformatics and Computational Biology Program, University of Idaho, Moscow, ID 83843, USA

³Swedish Environmental Epidemiology Centre, Department of Aquatic Sciences and Assessment, Swedish University of Agricultural Sciences, Uppsala SE75007, Sweden

⁴Faculty of Computer Science, University of Białystok, 15-351 Białystok, Poland

⁵Interdisciplinary Centre for Mathematical and Computational Modelling, University of Warsaw, 03-046 Warsaw, Poland

⁶Institute of Molecular Biology and Genetics of National Academy of Sciences of Ukraine, 03143 Kyiv, Ukraine

⁷Kyiv Academic University, 03142 Kyiv, Ukraine

*Correspondence address. Paweł P. Łabaj, Małopolska Centre of Biotechnology, Jagiellonian University, 30-387 Krakow, Poland. E-mail: pawel.labaj@uj.edu.pl

Abstract

Background: Microbiome-based disease prediction has significant potential as an early, noninvasive marker of multiple health conditions linked to dysbiosis of the human gut microbiota, thanks in part to decreasing sequencing and analysis costs. Microbiome health indices and other computational tools currently proposed in the field often are based on a microbiome's species richness and are completely reliant on taxonomic classification. A resurgent interest in a metabolism-centric, ecological approach has led to an increased understanding of microbiome metabolic and phenotypic complexity, revealing substantial restrictions of taxonomy-reliant approaches.

Findings: In this study, we introduce a new metagenomic health index developed as an answer to recent developments in microbiome definitions, in an effort to distinguish between healthy and unhealthy microbiomes, here in focus, inflammatory bowel disease (IBD). The novelty of our approach is a shift from a traditional Linnean phylogenetic classification toward a more holistic consideration of the metabolic functional potential underlining ecological interactions between species. Based on well-explored data cohorts, we compare our method and its performance with the most comprehensive indices to date, the taxonomy-based Gut Microbiome Health Index (GMHI), and the high-dimensional principal component analysis (hiPCA) methods, as well as to the standard taxon- and function-based Shannon entropy scoring. After demonstrating better performance on the initially targeted IBD cohorts, in comparison with other methods, we retrain our index on an additional 27 datasets obtained from different clinical conditions and validate our index's ability to distinguish between healthy and disease states using a variety of complementary benchmarking approaches. Finally, we demonstrate its superiority over the GMHI and the hiPCA on a longitudinal COVID-19 cohort and highlight the distinct robustness of our method to sequencing depth.

Conclusions: Overall, we emphasize the potential of this metagenomic approach and advocate a shift toward functional approaches to better understand and assess microbiome health as well as provide directions for future index enhancements. Our method, q2-predict-dysbiosis (Q2PD), is freely available (<https://github.com/Kizielins/q2-predict-dysbiosis>).

Keywords: microbiome, gut microbiome health, health index, dysbiosis, metagenomics

Introduction

The prevalence of a range of diseases and conditions peripherally or directly linked to microbiome health, such as inflammatory bowel disease (IBD), diabetes, obesity, and even various cancers, continue to increase globally, and substantial funds are currently spent on diagnosis and treatment [1, 2]. While a correlation between gut microbiome composition and human health is widely acknowledged [3], the accurate identification of microbial and host markers of disease states remains elusive. Accordingly, the ability to evaluate patient health status based on a gut microbiome snapshot would be of high clinical value. Stool-based methods are promising because they can be collected noninvasively and frequently, and analysis time is short. Furthermore, decreasing costs of stool sample analysis via next-generation sequencing makes such microbiome characterization a strong competitor as a diagnostic tool [4].

Dysbiosis, defined as a perturbation of gut homeostasis, is believed to be accompanied by reduced microbiota diversity and increased prevalence of “harmful” bacteria in adults [5, 6]. Eubiosis (opposite of dysbiosis) can be perturbed by a wide range of factors, including infection, diet, exercise, antibiotics, stress, or poor sleep [7]. The simplest interventions currently applied for the prevention or alleviation of mild microbiome dysbioses include dietary modification or prebiotics (often nondigestible food types that promote the growth of beneficial microorganisms), ingested live bacteria or probiotics (beneficial bacteria usually in capsules), and lifestyle changes. More severe cases of gut dysbiosis, failing to respond to the above interventions, may qualify for fecal microbiota transplants (FMTs), which are increasingly gaining traction in clinics worldwide [8]. However, host–microbiota and intra–microbiota interactions are both extremely complex and highly individual, and despite success with

Received: July 29, 2024. Revised: December 4, 2024. Accepted: February 5, 2025

© The Author(s) 2025. Published by Oxford University Press GigaScience. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

FMTs, we have as yet no real understanding of why or how they work.

There are a number of accepted approaches used for the evaluation of a given gut microbiome's health status based on stool composition. Alpha diversity (Shannon entropy, for example) is a frequent choice, as microbiome richness was long believed to be a key driver of microbiome health and robustness [9, 10]. Beta diversity has also been applied in a number of longitudinal studies, albeit to a lesser degree and mainly to identify eubiotic samples based on a time-resolved proximity to other healthy samples [11]. The most robust index to date, outperforming diversity indices, is the Gut Microbiome Health Index, or the **GMHI** [12], recently renamed the Gut Microbiome Wellness Index (**GMWI**). An updated version of this index has recently been published [13]. The **GMHI** is based on the ratio of 50 microbial species associated with healthy or unhealthy gut ecosystems and is reported to exceed 73% accuracy in determining disease state; thus, the authors suggest that gut taxonomic signatures can predict health status. Another metagenomic gut health index expanding on the **GMHI** approach, high-dimensional principal component analysis (**hiPCA**) [14], was introduced as a monitoring framework for personalized health purposes. The personalized approach is achieved by analyzing the contribution of each bacterium to the index, which allows for the identification of high-influence (ostensibly keystone) species in different patient groups. The **hiPCA** claim of better performance than the **GMHI** is attributed to the authors' application of additional transformation and clustering algorithms. Importantly, such studies are often defined by datasets limited in scope to industrialized nations and thus a less than complete consideration of diet–environment–microbiome interactions.

However, a recently revisited definition of the microbiome emphasizes the importance of not just the microbiota (a community of microorganisms) but the whole “theater of activity,” ToA [15]. This ToA includes structural elements (proteins, lipids, polysaccharides), metabolites, and environmental conditions [16]. It is tightly bound to its corresponding ecological niche, and the synergistic relations between species provide all the necessary, community-defining components. Based on this definition, we maintain that an index constructed from taxonomy alone is hardly sufficient to accurately capture biological phenomena occurring within the gut environment—the key to understanding gut dysbiosis. Instead, we hypothesize that **to effectively determine health**, (i) a metagenomic functional profile is required (microbiome phenotype) and (ii) species interactions (e.g., measured as co-occurrence) but not just presence should be considered.

We introduce an approach that is based on identifiable metagenomic features within ecosystems that extend beyond diversity measures and basic taxonomic information. This **function-centrism** is broached in 2 ways: (i) directly by evaluating the functional potential within and between species and (ii) indirectly by assessing co-occurrence and synergism between bacterial species. Our goal is not only to distinguish between healthy and diseased but also, importantly, to quantify the degree of dysbiosis in each sample for the given cohort. We derive the health-describing features based on an exploratory analysis of healthy samples from the Human Microbiome Project 2 [17] and outperform **Shannon entropy**, the **GMHI**, and the **hiPCA** in healthy versus IBD and obese classifications. The robustness of our index is further validated by corroboratively classifying 2 additional IBD-focused cohorts. By retraining the IBD-specific parameters on an additional set of 30 diverse cohorts encompassing a range of diseases, we demonstrate the superior performance of our approach. Our findings reveal that function- rather than taxonomy-based features

are more informative for the accurate classification of biological samples. Additionally, our method effectively identifies longitudinal microbiome changes in patients with COVID-19, which the **GMHI** and the **hiPCA** are unable to capture, and crucially, it is distinctly robust to sequencing depth. Our method, **q2-predict-dysbiosis (Q2PD)**, is freely available [18].

Results

High prevalence of “core functions” in health

In order to develop a strategy to assess the degree of dysbiosis in a given microbiome sample, we must define eubiosis (i.e., the healthy microbiome). We based our initial analysis (described in the Methods section) on 384 medically determined healthy samples from the HMP2 project and identified the most prevalent species, regardless of abundance (Fig. 1A). We observed that 50% of species were present in less than 5% of samples, and hardly any species were shared by all individuals. On the other hand, the prevalence of functions within the healthy population had an opposite trend—50% of functions were already represented by at least 40% of individuals (Fig. 1B), a functional redundancy unaccounted for in the **GMHI** (or the **hiPCA**, which is based on it). These results convinced us further about the unsuitability of basing an index purely on the presence of “core taxa” and encouraged a shift of focus toward more prevalent functions instead.

The addition of healthy samples from 2 validation cohorts maintained the function distribution profile obtained solely with HMP2 samples (Fig. 1C). In order to test whether the functions were universal or cohort specific, we calculated separately the distributions for functions present in 1, 2, or all 3 cohorts. We found that all functions missing from at least 1 cohort were present in less than 10% of samples, which indicated the presence of high-prevalence functions in all 3 cohorts. Based on the increased occurrence of certain functions in over 80% of samples (dotted line in Fig. 1C), we defined them as “core functions” (refer to <https://gigadb.org/dataset/102656> Supplementary Table S1 for full list). According to MetaCyc classification, 73.5% of “core functions” were assigned as “Biosynthesis” pathways, 18.8% “Degradation/Utilization/Assimilation,” and 7.6% “Generation of Precursor Metabolites and Energy.” In addition, a few of the above were additionally classified as “Superpathways.” The classification aligned well with a previously reported high prevalence of carbohydrate and amino acid metabolism-related pathways, potentially forming the functional microbiome core [19]. A detailed analysis of the core functions identified in our study, however, is out of the scope of this article.

Shannon entropy calculated on species and functions allowed for good discrimination between healthy and unhealthy samples in the HMP2 dataset (Fig. 2A, top). However, the trends were unclear in the 2 validation cohorts (Fig. 2A, middle and bottom), reflecting the need for more complex methodology, expanding beyond microbiome richness, in order to classify datasets without obvious separation. Diversity analyses revealed that the number of functions per sample remained similar or even increased during microbiome transitions from healthy state to dysbiosis in HMP2 (<https://gigadb.org/dataset/102656> Supplementary Fig. S1). While admittedly, this cannot be measured solely using metagenomics data, the similarity could hypothetically be due to altered expression of genes usually silenced in the eubiotic state, although this observation was not reproduced in the validation cohorts. Next, we investigated the presence of “core functions” in different groups, testing whether “core functions” are maintained

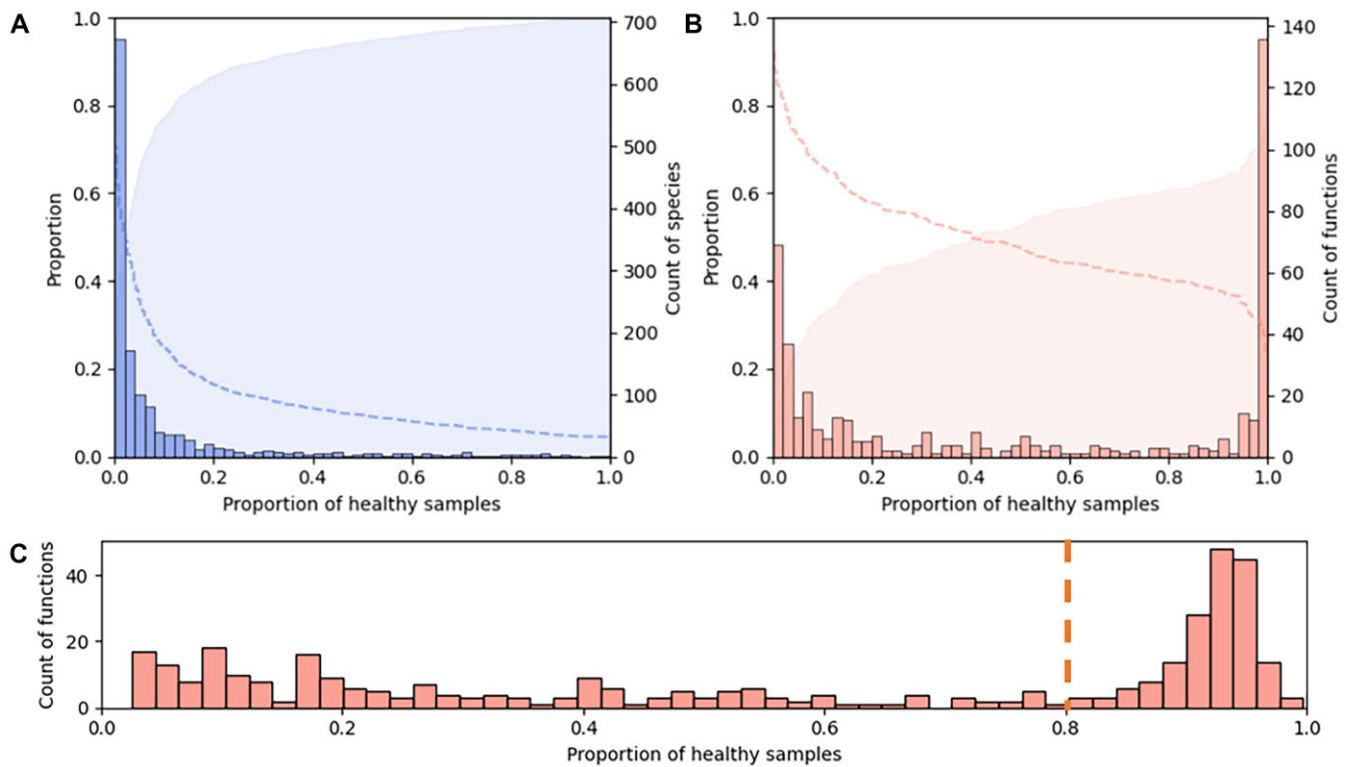


Figure 1: Distributions of species (A) and functions (B) present in healthy samples from the HMP2; absolute values of species and function counts are shown as histograms (with scales on the right-hand side, with shaded cumulative sum in the background and an inverse of the cumulative sum represented with a dashed line, with scales on the left-hand side). (C) Distribution of functions in healthy individuals from the HMP2 and 2 validation cohorts.

or replaced by others in dysbiotic, IBD samples. While the differences were not significant in most cases, we noted a visibly higher percentage of “core functions” and a higher percentage of all “core functions” in healthy as compared to disease samples (<https://gigadb.org/dataset/102656> Supplementary Fig. S2). We carried out differential enrichment analysis using Linear discriminant analysis Effect Size or LefSe [20], performed separately for each cohort, and identified 100 functions that were more abundant in healthy as compared to disease cohorts (Fig. 2B, <https://gigadb.org/dataset/102656> Supplementary Fig. 2). Over 90% of functions enriched in healthy samples were “core functions,” while they constituted less than 5% of functions enriched in the unhealthy class of validation cohort 2 and HMP2 (Fig. 2C). The LefSe analysis on validation cohort 1 revealed only 10 significantly enriched functions (7 in health and 3 in disease), all of which were core. This indicated a more heterogeneous functional landscape within this cohort.

Species interactions and function contributions in health

Corroborating results from past studies, we observed a decrease in the species abundance in dysbiotic samples (<https://gigadb.org/dataset/102656> Supplementary Fig. S3, [21, 22]). Having previously noted an increase in the number of functions (<https://gigadb.org/dataset/102656> Supplementary Fig. S1), we speculated that the remaining species may contribute to core or new functions, forming new connections with one another. Due to the substantial number of initial connections to analyze (170 “core functions” and 1,490 species present in at least 2 projects), we restricted the number of species to those most informative in the context of health/disease

state separation. We chose the Multi-Dimensional Feature Selection (MDFS) algorithm, as it was the only feature selection method accounting for interfeature interactions that we were aware of at the time of manuscript submission [23, 24]. This approach reduced the number of relevant species to 587, allowing us to eliminate noise and focus on the most important interactions (see Methods for more details about the feature selection procedure).

We then used the SparCC algorithm, designed specifically for compositional data, to investigate correlations between the MDFS-selected species in health and disease [25]. We did not observe any trends in the number of correlations or in the fraction of positive correlations per group that would indicate differences between the two. However, we identified opposite relationships of some species in different groups (Fig. 3A). A number of species we found to be positively correlated in eubiosis and are generally considered beneficial (e.g., *Eubacterium rectale*, *Faecalibacterium prausnitzii*, and a number of *Bacteroides* species), and those relationships would be disrupted in dysbiotic groups. We observed that the prevalence of the pairs positively correlated with health was higher than in a number of disease-associated groups (Fig. 3B). Due to this, we included the co-occurrence of such species as another feature of interest to aid potentially in the determination of microbiome health.

Based on our previous results, we hypothesized that the contributions of each species to functions would be relatively stable in the healthy state and less predictable in disease. To test this, we compared the contributions of MDFS-identified species to “core functions” in different groups (<https://gigadb.org/dataset/102656> Supplementary Fig. S4). We did not observe any differences between health and disease, despite a relatively tight clustering of the healthy groups. However, we found stronger results

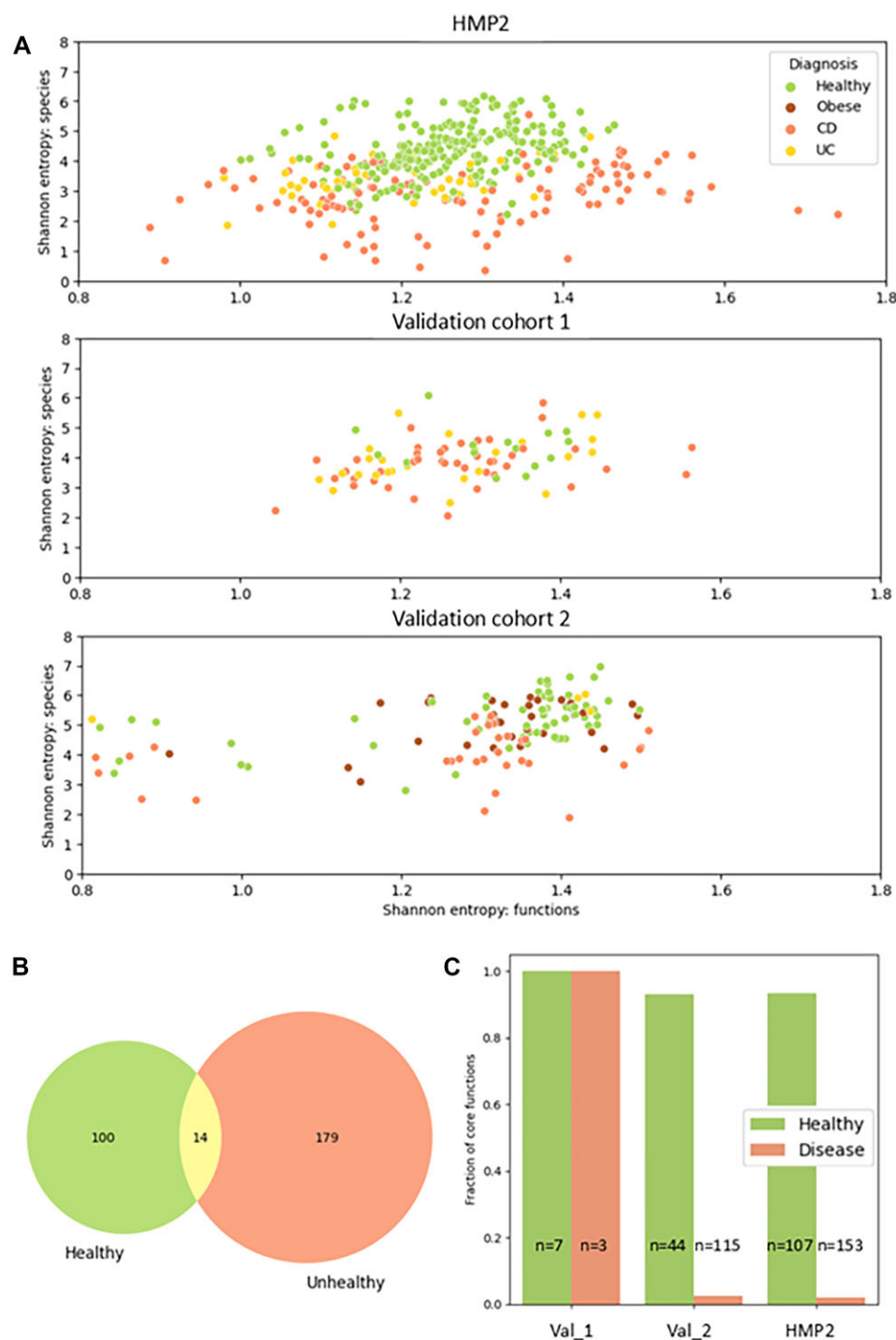


Figure 2: (A) Shannon entropy scores for species and functions in healthy and unhealthy samples from the HMP2 and validation cohorts. (B) LEfSe differential enrichment analysis: overlap of enriched pathways in healthy and unhealthy individuals in the validation and HMP2 projects. (C) Fraction of core functions among differentially enriched functions in healthy and unhealthy individuals in the validation and HMP2 cohorts.

when exploring functional redundancy. While the average number of species per function and the average number of functions per species did not reliably separate healthy from diseased profiles (<https://gigadb.org/dataset/102656> Supplementary Fig S5), the latter approach was more informative, as described in detail below. This finding was congruent with our earlier suspicions of an inherent functional plasticity of microbiome structure, with mod-

ulation of function altering connectivity in the interaction network, leading to a shift toward less abundant, noncore functions upon perturbation of homeostasis. It also highlighted the challenge of identifying dysbiosis based on singular features, which were never statistically significant for all cohorts, and the utility of multiple perspectives for microbiome description, in order to meaningfully classify them.

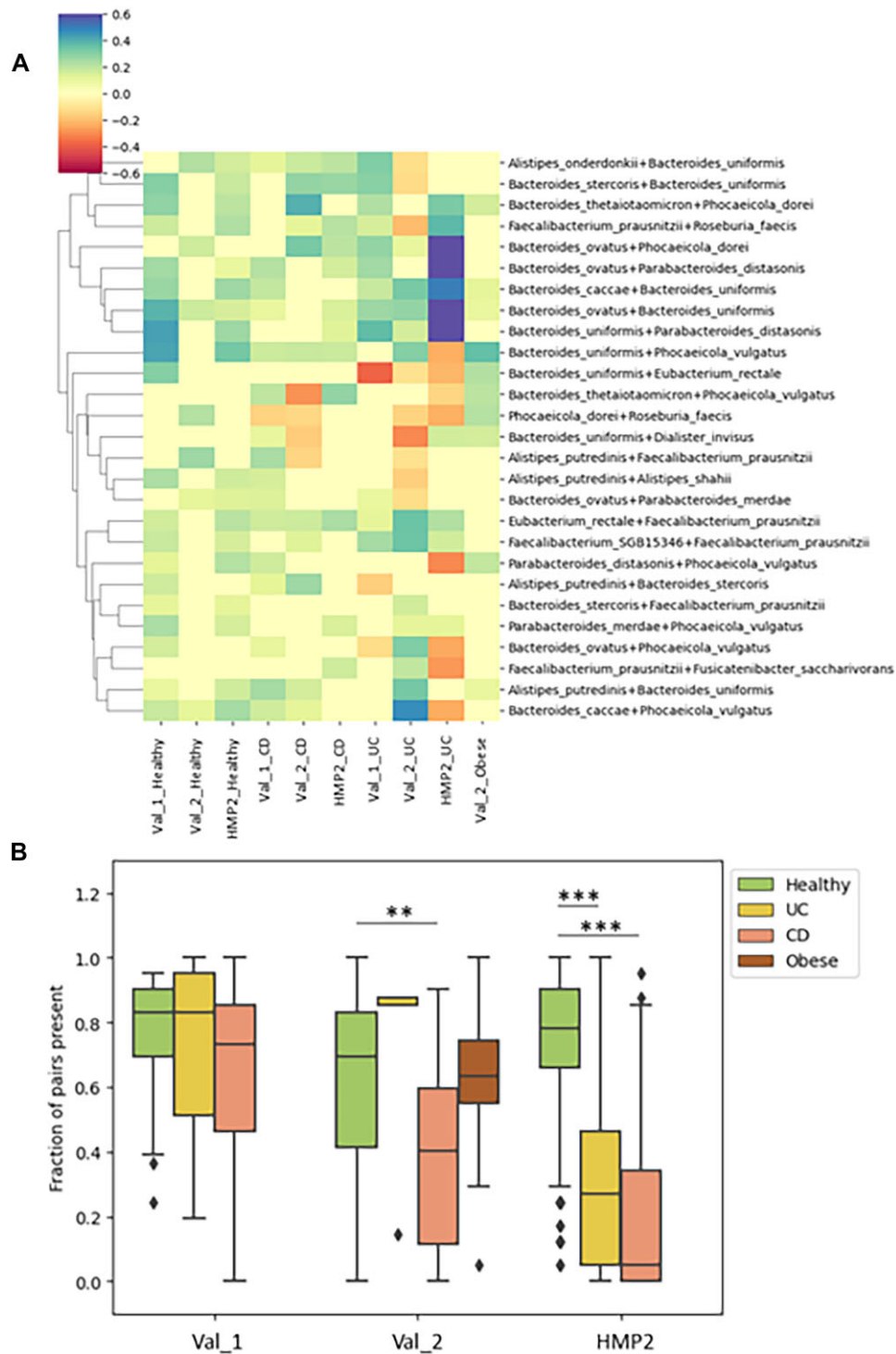


Figure 3: (A) SparCC correlation strengths between species, restricted to pairs that were not negatively correlated in any healthy cohort. (B) Prevalence of the pairs in different cohorts.

Testing the accuracy of prediction for healthy and IBD individuals

Our final set of health-defining microbiome features included the following parameters (details of how each feature was calculated can be found in Table 1):

- i) the fraction of “core functions” found,
- ii) the proportion of “core functions” among all functions,

- iii) the proportion of co-occurring species pairs in healthy samples, and
- iv) the average number of functional “contributions” per species.

In addition, we included 2 parameters derived from the GMHI method—the number of “good” and “bad” GMHI species identified in a sample, which would enable us to compare between approaches. We then fed these described parameters into a machine

Table 1: Parameters of the Q2PD model

Name	Explanation	Derivation
Frac_of_core_functions_found	Fraction of “core functions” found	Number of “core functions” in a sample/number of all “core functions” in our list
Frac_of_core_functions_among_all	Fraction of “core functions” among all functions	Number of “core functions” in a sample/number of all functions in a sample
Species_found_together	Fraction of species pairs commonly occurring together in healthy samples	SparCC correlation of species >0.1, only nonnegative correlations in healthy cohorts
Func_contributions_per_species	Average number of function contributions per species	Number of all species to functions contributions based on stratified output/number of all species in a sample
GMHI_good	Number of “good” GMHI species	List of health-associated GMHI species
GMHI_bad	Number of “bad” GMHI species	List of disease-associated GMHI species

learning model. We opted for a random forest classification algorithm [26] due to its robustness to imbalanced data and interpretability and performed a leave-one-out cross-validation [26]. For the model training, validation, and subsequent testing, we used taxonomic and functional profiles from the curated Metagenomics database [27] to ensure consistency and reproducibility (see Methods).

Our index demonstrated the strongest statistically significant separation between healthy and Crohn’s disease or ulcerative colitis individuals across all methods evaluated (Fig. 4A). Alongside hiPCA, it was 1 of only 2 approaches to achieve a statistically significant distinction in the Nielsen_2014 cohort. Overall, both our index and hiPCA exhibited comparable levels of accuracy and AUC across the 3 cohorts, outperforming the other methods by a notable margin (Fig. 4B). By contrast, the GMHI consistently delivered mediocre results, and Shannon entropy only performed well applied to the HMP2 cohort, yielding low area under the curve (AUC) scores (0.37–0.53) in all other cases. When index values were used as features in the Boruta algorithm [28], with health status (0/1) as the target variable, our index emerged with the highest mean and summed importance across all cohorts (Fig. 4C) [28]. The above highlights the ability of our index to provide the most informative scores for health status prediction as compared to the other methods.

Beyond IBD

While only healthy and IBD individuals had erstwhile been included in the development and validation of our approach, we wondered about the applicability of the Q2PD to dysbiosis attributed to other diseases. We extended our dataset to another 27 additional cohorts from various disease states, equating to 30 datasets used for method validation. We did not change any parameters of the Q2PD, which were originally determined based on the healthy samples from the HMP2. Instead, we retrained the model with the new data and performed a leave-one-cohort-out approach to ensure a robust benchmark. The procedure placed our index at a disadvantaged position, as some of the added datasets had been used to develop and train the other methods, and our approach was thus truly blind to outcome in these new cases.

Despite the disability, and gratifyingly, the Q2PD achieved in terms of performance the highest average accuracy and AUC across all datasets {(AUC = 0.61, accuracy = 0.58) > hiPCA (AUC = 0.58, accuracy = 0.57) > GMHI (AUC = 0.55, accuracy = 0.55) > Shannon entropy on species (AUC = 0.52, accuracy = 0.53) >

Shannon entropy on functions (AUC = 0.44, accuracy = 0.43)} (see Fig. 5A). The consistently poor performance of both entropy-based measures suggested their highly limited utility as predictive indices.

We observed that where Q2PD classified a particular cohort better than the other indices, it did so with a significantly greater margin than when it lost to the other methods. Its average winning AUC margin over the hiPCA, the second best classifier, was 0.19 while the losing margin to the hiPCA was 0.09 when the hiPCA had the highest AUC. The average winning AUC margin of Q2PD against the mean of the other indices was 0.20 and 0.11 if any other was better. In both cases, the t-test statistics for the differences between the means of the Q2PD’s winning AUC margins and that of the hiPCA or the average of others produced P values of 0.03 and 0.02, respectively, indicating a significant classification improvement with our method in areas in which the remaining indices did not classify well. The improvement was even more striking when we excluded datasets that had been used for the training of either method. In this case, the winning margin of the Q2PD was 0.22 while the losing margin to the mean of the other indices if any of them was better was 0.09.

Our investigation into the accuracy and AUC of the indices for each cohort revealed substantial variability in terms of the classes of cohort that each index was able to classify (<https://gigadb.org/dataset/102656> Supplementary Fig. S6). We observed that while some cohorts such as Liss_2016 could be classified well with function-based indices (Q2PD and Shannon entropy on functions), other cohorts such as Gupta_2019 were slightly better separated with taxonomy-based indices (hiPCA, GMHI, and Shannon entropy on species). An exploration of the importance of feature(s) associated with the model training on each dataset alone revealed a large amount of diversity, suggesting different kinds of information used to classify different cohorts (Fig. 5B). Interestingly, when trained on the level of individual cohorts, a random forest would pick the “GMHI_bad” as its most informative parameter (scoring the lowest rank in 9 cases) and the “GMHI_good” as the worst (appearing at the bottom of the ranking 6 times). The importance of the function-based features would vary depending on the dataset, with “Contributions_per_species” winning and losing twice and the other two being consistently in the middle of the ranking.

Having observed a discrepancy between the superior Q2PD performance and the greatest importance of the “GMHI_bad” parameter, we investigated this further by testing which types of diseases were best suited for each index to utilize. To this

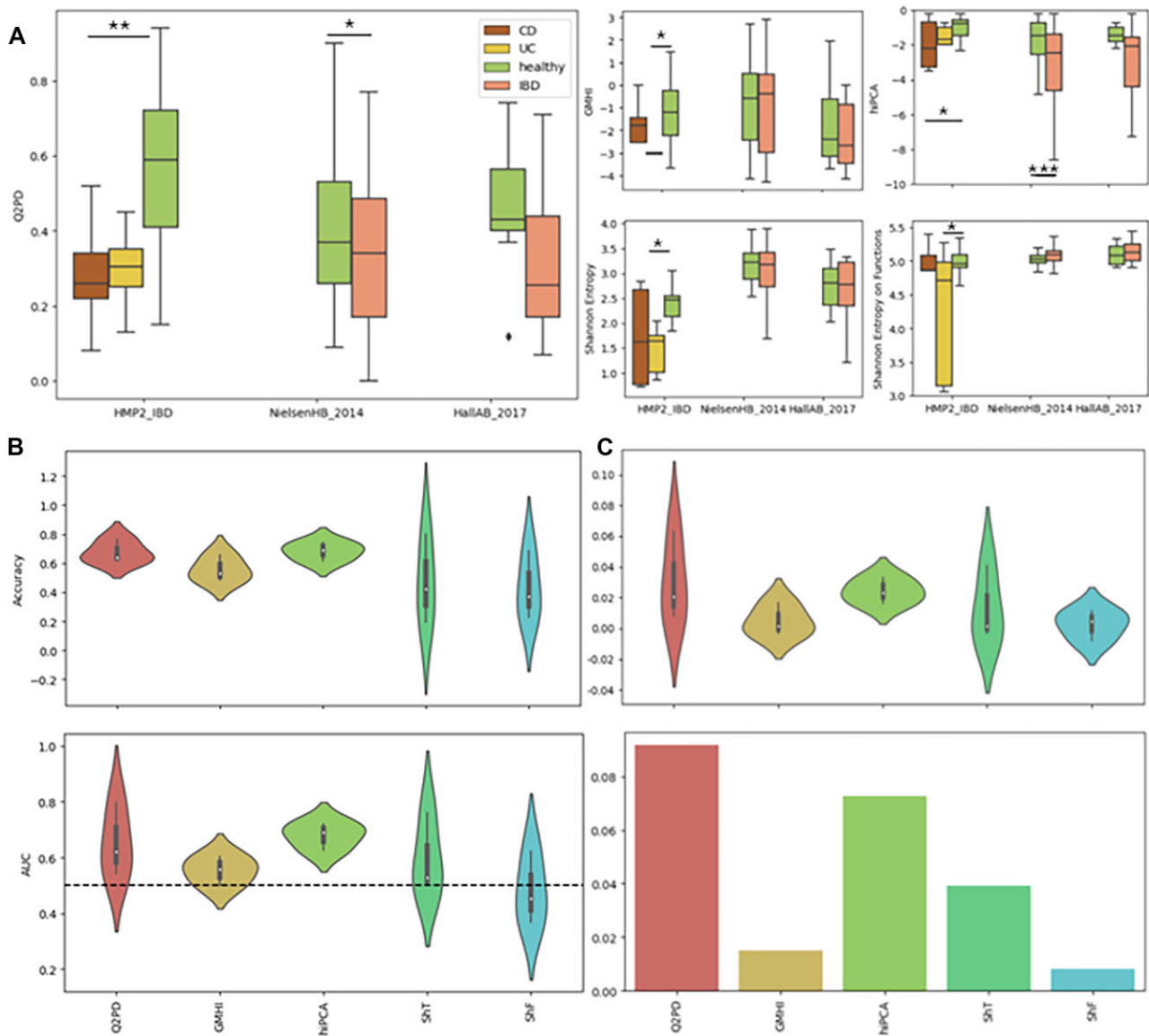


Figure 4: (A) Q2PD, GMHI, hiPCA, and Shannon entropy (on species and functions) scores for healthy and IBD (CD: Crohn's disease; UC: ulcerative colitis) individuals. (B) Accuracy and AUC values for each index, per IBD cohort. ShF: Shannon entropy on functions; ShT: Shannon entropy on taxa. (C) Average (top) and summed (bottom) importance of each index in the context of IBD prediction, determined by Boruta.

end, we designed a ranking method based on the number of times each index achieved the highest AUC for the most cohorts for each disease. Overall, Q2PD outperformed the other indices for 6 diseases (atherosclerotic cardiovascular disease, colon cancer, infection, metabolic disease, schizophrenia, and fecal microbiota transplant: donor versus patient classification). This was followed by hiPCA for 5 (Behçet's disease, IBD, type 2 diabetes, chronic fatigue and cirrhosis), Shannon entropy on taxa for 2 (Parkinson's disease and acute diarrhea), and GMHI for 1 (a cohort with mixed diseases). From this, we also calculated an average rank for each method for all the diseases above, and yet again, Q2PD ranked best (average ranking = 2.39), followed by hiPCA (2.71), Shannon entropy on taxa (3.14), GMHI (3.32), and Shannon entropy on functions (3.57). The poor performance of the GMHI indicated a diminished role of the "GMHI_bad" parameter when combining all datasets and implied a better generalization of the health status using function-based parameters.

Q2PD robustness to longitudinal alterations and sequencing depth

To construct a final model, we trained the random forest classifier on the complete set of 30 cohorts. We then took advantage of a longitudinal COVID-19 dataset that had been sequenced both shallowly and deeply, and was "unseen" by any methods, in order to evaluate performance of Q2PD across sequencing depths. The set consisted of 3 groups—COVID-19 patients who, during the course of the treatment, were either (i) transferred to the intensive care unit (ICU) or (ii) recovered (noICU), and (iii) controls (healthy hospital staff). For every individual, 2 time points were selected—"1," which was collected upon hospital admission or, in the case of staff, early in the pandemic, and "2," which was the final sample taken from each individual. As expected, sequencing data with fewer than 300,000 reads per sample fell short in accurately classifying individuals, primarily because the limited read depth did not provide sufficient coverage for comprehensive functional annotation (Fig. 6A).

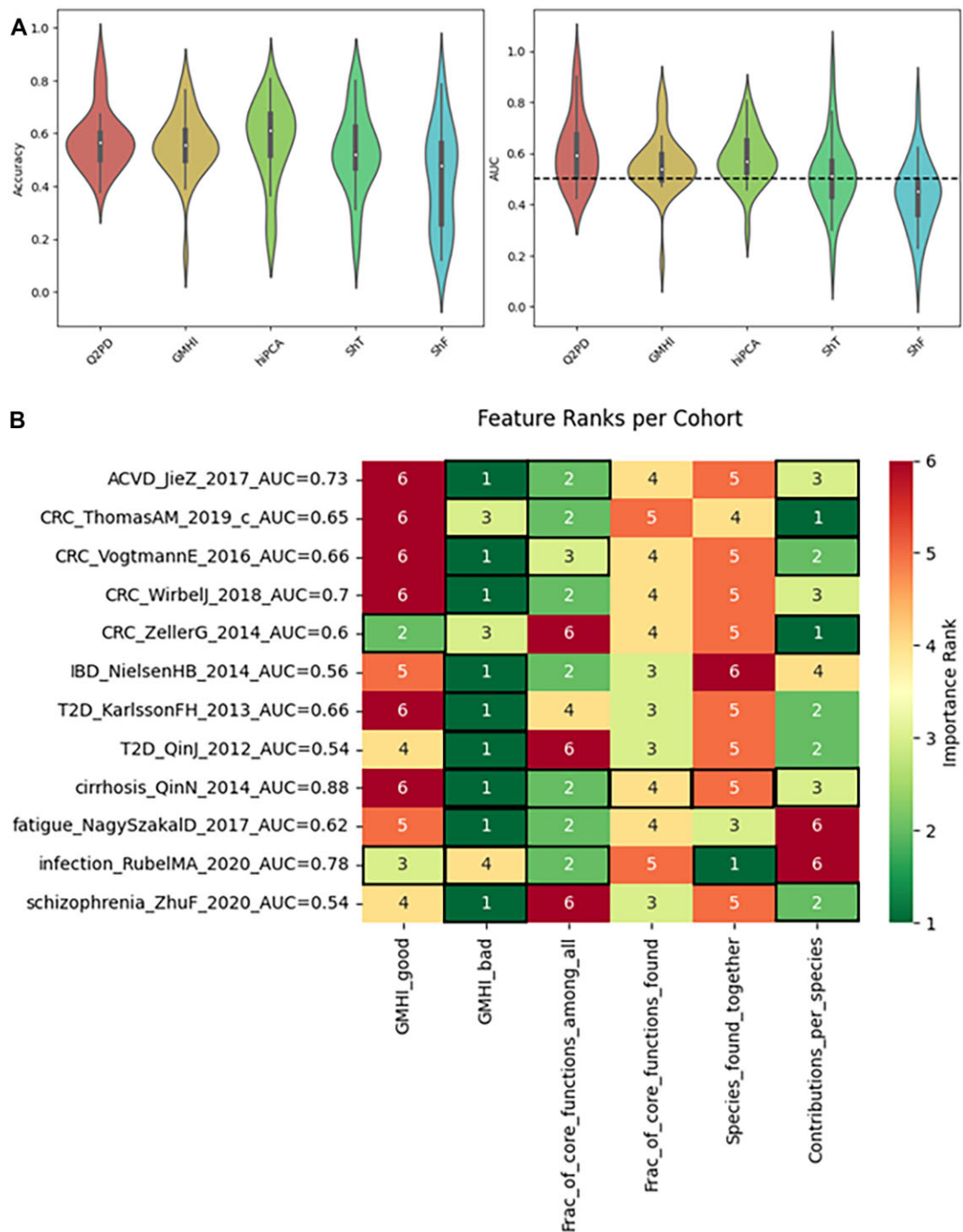


Figure 5 (A) Average accuracy and AUC values for each cohort. (B) Average importance ranks of features for each cohort. A lower rank indicates greater importance. Cells with black borders indicate variables identified as informative by the Boruta algorithm. The AUCs were produced as a part of the feature importance analysis when training on individual cohorts and therefore do not align with the AUCs produced by the Q2PD. Only datasets with AUCs >0.5 are shown.

Surprisingly, despite its earlier described successes at classification, Q2PD was not able to distinguish between healthy and COVID-19 individuals sequenced deeply classifying ICU patients as healthy (Fig. 6B). Curious about the reasons for this, we investigated feature importance and discovered that the 2 taxonomic parameters “GMHI_good” and “GMHI_bad” both had negative values. This was indicative of the detrimental influence of taxonomy-based features on the performance of the model. As suspected, retraining the Q2PD without them led to the expected, correct

predictions (Fig. 6C), suggesting that this dataset could be classified based on functional information alone. We further validated this when we performed classification using the taxonomy-based GMHI (Fig. 6D) and the hPCA (Fig. 7E), which was inaccurate for GMHI and opposite for hPCA.

Q2PD performed poorly on the shallow COVID-19 dataset, which led us to ask where the threshold lay for sequencing depth reliability. In order to investigate method robustness to depth, we applied various degrees of rarefaction to deeply sequenced

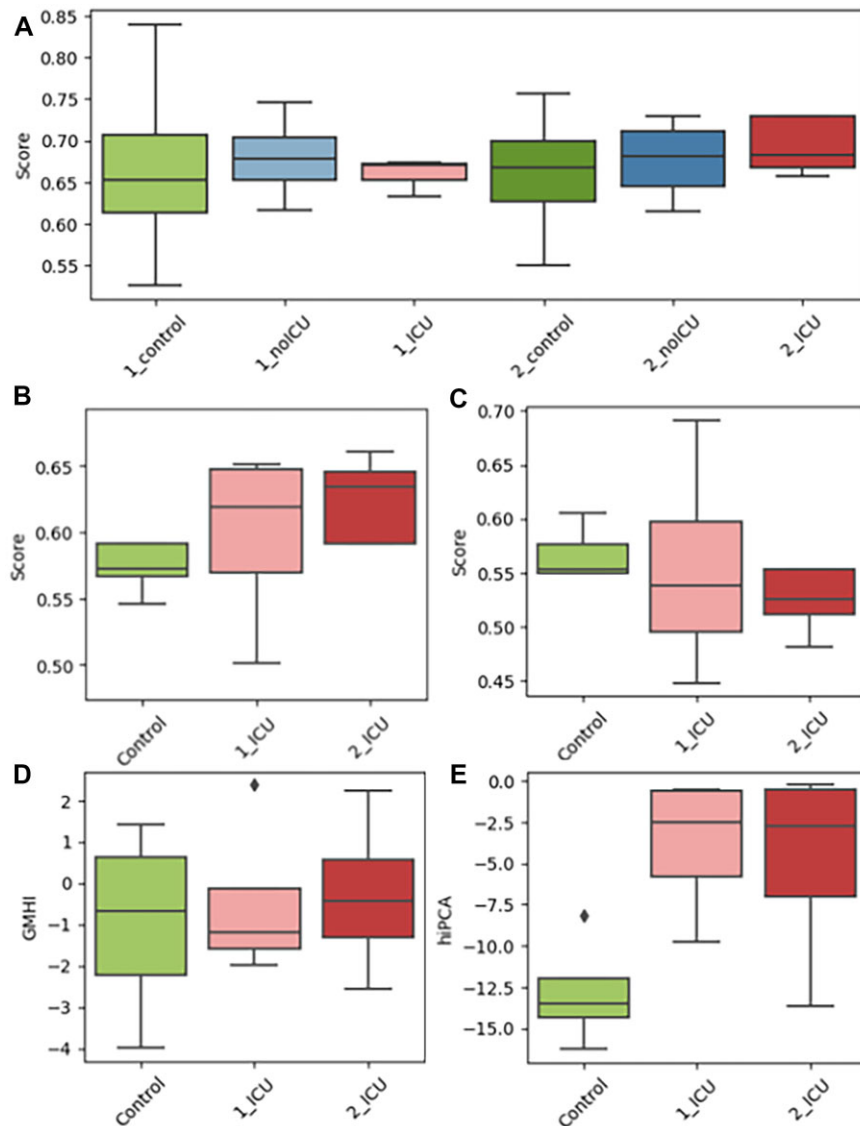


Figure 6: (A) Q2PD predictions on the shallow COVID-19 cohort. (B) Q2PD predictions on the deep COVID-19 cohort. (C) Q2PD predictions on the deep COVID-19 cohort with the taxonomic parameters excluded. GMHI (D) and hiPCA (E) predictions on the deep COVID-19 cohort.

control samples, expecting to see similar prediction scores despite variable sequencing depth. Because the performance of the **hiPCA** (second best classifier after the Q2PD overall) on the COVID-19 cohort was worse than that of the **GMHI**, we decided to use the latter as our benchmark.

The scores produced by the Q2PD were robust and consistent regardless of sequencing depth, whereas for **GMHI**, they increased with degree of rarefaction (Fig. 7A, B). Furthermore, the **GMHI** classification appeared to be only relevant for a foreboding depth of 2 million reads (score >0), as the healthy samples were defined as unhealthy at greater sequencing depth (score <0). Our index correctly identified healthy samples at any sequencing depth (scores were always above threshold).

We plotted the values for each feature separately and found that at low sequencing depths, certain species (Fig. 7C, D) or “core functions” (Fig. 7E, F) were undetectable. The “Fraction of core features among other features” and the “Function contributions per species” were those most dependent on coverage, which was anticipated due to their dependence on low-abundance functions (Fig. 7E–H). Interestingly, the numbers of “good” and “bad” species

identified at any sequencing depth covered only 32% of the complete **GMHI** list (2 and 14 versus 7 and 43 for good and bad species, respectively). In line with the underlying theme of our work and in agreement with all other presented data, this overlap argues strongly against indices that are based solely on taxonomy.

Discussion

A connection between the human gut microbiome and gut health is now well established, and a number of approaches have been employed in an effort to identify gut dysbiosis from sequence-based analysis of stool samples. Those methods are based on taxonomy and rely either on measures of microbiome richness (alpha/beta diversity) or on the presence or absence of so-called “good” and “bad” bacteria, with the health-indexes **GMHI** [12] and **hiPCA** [14] proposed formally. Hampering such approaches, however, are ecological considerations of metabolic or functional redundancies inherent within complex environments. Currently, redefining the microbiome to include **inferred functionality** is bolstered by recent studies that highlight the importance of

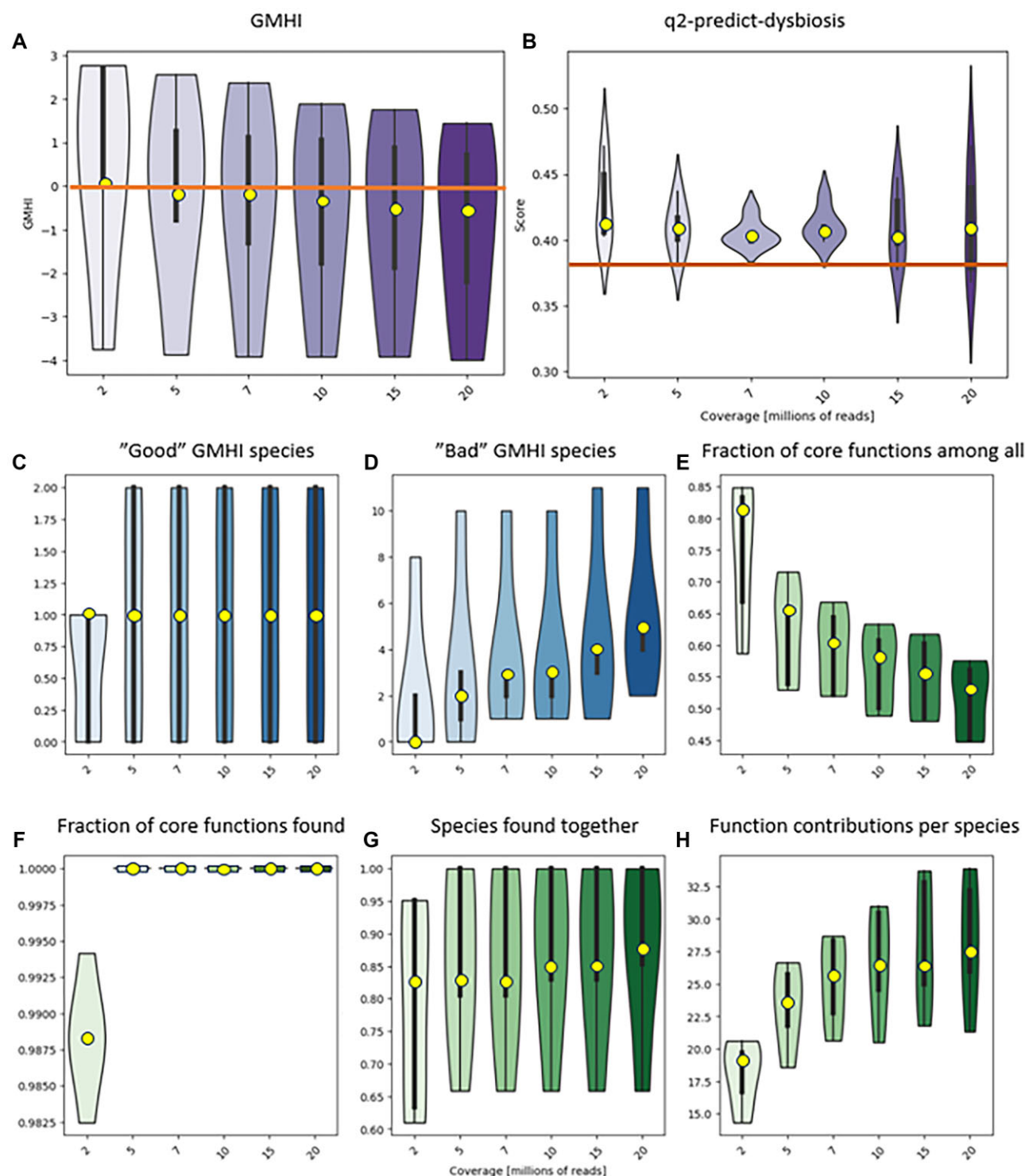


Figure 7: Robustness of health predictions to sequencing coverage. **GMHI** (A) and **Q2PD** (B) scores for deeply sequenced healthy samples from the COVID-19 cohort, rarefied to corresponding depths. The horizontal lines represent health thresholds relevant to the corresponding methods. Taxonomic **GMHI**-inspired **Q2PD** features (C, D) remain relatively stable above the coverage of 5 million reads. Functional and species interaction-related features are much more sensitive (E–H).

interactions between microbiome components and functional aspects thereof. To address the inadequacy of the dogmatic Linéan (taxonomy-based) approach for evaluating change in microbiomes, we developed a novel method that incorporates function in bioinformatics-based assessment of microbiome dysbioses. We show that features based on microbiome functions and interac-

tions define a healthy microbiome more accurately. There exists a set of “core functions,” which are consistently identified as present in healthy gut microbiomes and disappear in the advent of dysbiosis. We compare our results to the **hiPCA**, the **GMHI**, and 2 **Shannon entropy** measures (calculated on either species or on functions). Our index outperforms the other methods not only in

the originally targeted IBD classification but also when applied to a range of other diseases. Finally, it is robust to sequencing depth, unlike the **GMHI**, despite being based on sequencing depth-sensitive parameters.

While we here present the index **Q2PD**, we acknowledge that challenges remain prior to eventual clinical deployment as the method's robustness for diseases other than IBD and obesity is improved. One of the most fascinating results from this work, apart from the model itself, was the finding that different parameters were of varying importance across different diseases and cohorts. This finding could have practical implications in the clinic, as it could indicate the directionality of the microbiome-disease connections. By this, we mean specifically that diseases originating in the gut (e.g., IBD) are usually associated with taxonomic shifts and thus better classified with taxonomy-based indices while diseases originating elsewhere may have (functional) effects on the gut and thus better identified with function-oriented methods. It should be noted, though, that the etiology of many diseases (i.e., whether they originate in the gut or not) remains unknown.

Clearly, a deeper understanding of how the identified function- and interaction-based microbiome features respond to variations in sequencing depth and quality is required. Our initial test did find that longitudinal data provide more insight into personalized "core features" of a microbiome, and this may be indicative of individual deviations from the normal. Thus, tracking microbiome changes in individuals over time may facilitate the early identification of microbiome trajectories or alterations that could help determine risk for certain conditions or diseases, potentiating personalized clinical intervention.

A caveat to this type of study is the infancy of certain fields, in particular, the availability of functional annotation software. Currently, the only widely accepted functional annotation software is HUMAnN (any version), which requires data to be processed in a particular way. In future work, we plan to explore other possibilities and ultimately migrate to other inputs (such as mifaser) as well as toward a more universal solution, such as gene content [29]. Only then can recent developments in augmented functional annotation be efficiently utilized, as we already have shown in other applications [30–32].

We are also considering the introduction of a multiomics approach to improve our predictions, combining metagenomics, metabolomics, and proteomics data. With decreasing costs for each of the above, predicting patient health status based on all 3-omics methods might be a reasonable option that improves accuracy and prediction in the not-too-distant future. It would also allow us to quantitatively estimate the functions directly, instead of basing our analyses on the functional potential evaluated based on metagenomics data. As a primer for our current efforts, the integration of multiple levels could be performed using the aforementioned MDfS, by investigating relationships between the layers, or using more advanced network-based approaches such as ViLoN [33].

In conclusion, we highlight the performance and high accuracy obtained by **Q2PD**, based only on a limited number of highly relevant parameters. Despite its development to separate samples derived from healthy versus IBD individuals, its broad applicability was proven to accurately distinguish between healthy patients and individuals with other disease states and conditions. Motivated by this success, we are convinced that our approach and methodology provide a better mechanistic understanding of the human gut microbiome and its health, one based on resource competition and interactions based on fundamental ecological principles that ultimately define microbial community structure.

Methods

Defining metagenomic parameters associated with health

The raw shotgun sequencing fastq files for the HMP2 and the 2 IBD validation cohorts used to derive metagenomic health-associated parameters were processed with **Trim Galore 0.6.10** (RRID:SCR_011847) [34] to ensure sufficient read quality. The taxonomic profiles were calculated with **MetaPhlAn 4.0.6** (RRID:SCR_004915) [35] and functional profiles with **HUMAnN 3.7** [36]. Species selection was performed using MultiDimensional Feature Selection, or the **MDfS** [23, 24], with a Benjamini-Hochberg *P* value correction in a 2-dimensional mode. Selection was performed per cohort (healthy vs. each of the nonhealthy groups in separate **MDfS** runs), and the final list of species was the union of the MDfS results for all cohorts (corrected *P* < 0.05). Metagenomic table formatting and alpha diversity plots were done using **QIIME 2** [37] and **SparCC** correlations with the **SCNIC** plugin [38]. Plots were made using custom Python scripts. Any statistical tests were calculated with the independent *t*-test, unless stated otherwise.

IBD index

Random forest is a classification and regression method [26]. The algorithm uses an ensemble of CART trees [39], in which each tree is built using different bootstrap samples of data and different random subsets of variables at each stage of the tree construction. It is a robust and versatile algorithm that works well on different types of data [40].

Metagenomic features passed to the random forest were constructed based on healthy samples from the HMP2 and 2 validation cohorts (Val_1 and Val_2). The features are presented in Table 1.

The importance of the features was assessed using a permutation approach and 5-fold cross-validation.

From this step on, we used the manually curated samples from the curatedMetagenomicsRepository to ensure consistent sample processing (MetaPhlAn [35] and HUMAnN [36]; specifically, we used the MetaCyc functional annotations), as well as to provide an additional layer of validation of our approach. Our manual curation excluded a sample for the following reasons: incomplete metadata, use of antibiotics or other similar drugs, repeats from the same patients, young age (newborns), or healthy samples not being fully healthy (high body mass index, chronic conditions). During filtering, we set the following thresholds: species abundance $\geq 0.1\%$, pathway coverage $\geq 20\%$, and function abundance $\geq 0.01\%$ —with any features below that changed to 0.

During random forest training, healthy individuals were labeled as "1" and those with the IBD as "0." A leave-one-out cross-validation procedure was applied to predict health scores for each sample to avoid model overfitting or the need to split the data into training and testing sets. The final health score is the output of the random forest's *predict_proba* method, which expresses the probability of each sample belonging to the healthy (or unhealthy) group.

The GMHI scores were calculated with the **QIIME 2 q2-health-index** plugin [41], and the **hiPCA** [14] predictions were obtained by substituting the original test file with our test samples due to a lack of instructions about how to do it better (this could potentially result in an overlap of the train and test samples, possibly falsely increasing the **hiPCA**'s accuracy). The Shannon entropies were calculated separately on taxonomic and functional profiles for each sample using custom Python scripts.

The GMHI was the only parameter with a predefined threshold for health. In the case of the other methods, the optimal threshold would be defined using the Youden's *J* statistic calculated on the training data.

The relevance of each index in the context of IBD predictions was calculated using **Boruta** [28]. The Boruta algorithm is a wrapper around the random forest classifier. It works in the following way: a dataset is extended by adding a randomly permuted copy of each original variable—a so-called shadow variable. A predictive model for the decision variable is then built using the random forest algorithm. The importance of each variable is estimated using a permutation test. Boruta collects the information about the importance and then compares the importance score of each original variable with the maximal importance achieved by a shadow variable.

The procedure is repeated multiple times, and a statistical test is performed. The variables are eventually assigned to 3 classes: Confirmed (better than random), Rejected (no better than random), and Tentative (those that could not be assigned to the Confirmed or the Tentative class). The utility of the variable is a good indicator of the information importance carried by the variable and also in situations where the synergistic interactions are important.

In the Boruta section of our analysis, the predictions of each index were passed as parameters, and the health labels (0/1) were used as decision vectors. Index importances were redefined as ranks—with the highest importance marked as rank “1.”

Index testing: expanding to other diseases

For the purpose of testing, the index was retrained on the complete set of 30 datasets. However, no parameters were modified. To ensure a reliable benchmark, a leave-one-cohort-out procedure with a 5-fold cross-validation was performed. The accuracy and AUC statistics were defined separately for each test cohort based on the validation set (which was a subset of the training dataset).

The performance of our index was evaluated by calculating its winning margin over hiPCA, the second best index, and the mean of the other methods. Specifically, for each test dataset, the AUC of the hiPCA or the mean AUC of the other methods was subtracted from the AUC produced by the Q2PD. The winning margin was defined as a mean of the AUC differences, separately for cases when Q2PD won and when not.

Importance of the features contributing to the Q2PD score was evaluated by training a random forest model separately for each cohort and using a feature permutation approach. The importances were changed to ranks, with the highest ranks (lowest values) representing the greatest importance.

Final index and application to the COVID-19 dataset

The final index was trained on the complete set of 30 cohorts using 5-fold cross-validation. The health threshold was defined based on a mean of thresholds defined on training data for each iteration of the leave-one-cohort-out approach and was set at 0.38 (which was very similar for all iterations).

To evaluate the indices in the context of the robustness to sequencing depth, healthy deeply sequenced samples from the COVID-19 dataset were rarefied to different sequencing depths using **seqtk 1.4** [42].

Availability of Supporting Source Code and Requirements

Project name: q2-predict-dysbiosis

Project homepage: <https://github.com/Kizielins/q2-predict-dysbiosis>

Operating system(s): Platform independent

Programming language: Python

Other requirements: scikit-learn version $\geq 1.1.3$

License: MIT License

RRID:SCR_026038

bio.tools ID: q2-predict-dysbiosis

Additional Files

Supplementary Fig. S1. Number of observed functions per sample, separated by health group and project.

Supplementary Fig. S2. Comparison of core functions in health and disease. (a) Overlap of LEfSe differentially enriched pathways in healthy versus unhealthy samples in the validation and HMP2 cohorts. (b) Overlap of LEfSe differentially enriched pathways in unhealthy versus healthy samples in the validation and HMP2 cohorts. (c) Fraction of seen core functions among all annotated core functions. (d) Fraction of seen core functions among all functions seen in a sample.

Supplementary Fig. S3. Number of observed species per sample, separated by health group and project.

Supplementary Fig. S4. Top species contributions to core functions, based on the stratified output of HUMAnN. For every cohort, the sum of the contributions for all samples was calculated, and contributions with the greatest summed abundances are shown. The final values have been normalized to values between 0 and 1 to compare the strength of the correlations between the groups.

Supplementary Fig. S5. Functional redundancy. (a) Average number of species per function in a sample. (b) Average number of functions per species.

Supplementary Fig. S6. Accuracy and AUC scores achieved by each index, separately for each cohort.

Supplementary Table S1. “Core functions”: functions present in at least 80% of healthy individuals from the HMP2 and 2 validation cohorts.

Abbreviations

AUC: area under the curve; FMT: fecal microbiota transplant; GMHI: Gut Microbiome Health Index; GMWI: Gut Microbiome Wellness Index; hiPCA: high-dimensional principal component analysis; HMP2: Human Microbiome Project 2; IBD: inflammatory bowel disease; ICU: intensive care unit; MDFS: Multi-Dimensional Feature Selection; Q2PD: q2-predict-dysbiosis; ToA: theater of activity.

Acknowledgments

We thank Dr. Krzysztof Mnich, Dr. Sajad Shahbazi, Dr. Balakrishnan Subramanian, and Piotr Stomma from the University of Białystok for their expertise in the Multi-Dimensional Feature Selection algorithm and insightful contributions to our work; Dr. Tomasz Kościółek from the Sano Centre for Computational Medicine; and the members of Bioinformatics Research Group at MCB JU for their comments and support.

Funding

This research was conducted as a part of the NCN Sonata BIS grant number 2020/38/E/NZ2/00598. The open-access publication of this article was funded by the Priority Research Area BioS under the program “Excellence Initiative – Research University” at the Jagiellonian University in Krakow. We gratefully acknowledge Polish high-performance computing infrastructure PLGrid (HPC Center: ACK Cyfronet AGH) for providing computer facilities and support within computational grants no. PLG/2023/016234 and no. PLG/2024/017180.

Data Availability

The HMP2 data can be found in Sequence Read Archive (SRA) with the accession code PRJNA398089. The 2 IBD validation cohorts can be located under accessions PRJNA389280 and PRJEB1220. The test cohorts were downloaded from the curatedMetagenomics-Data repository [27]. The samples from this repository were manually selected—we removed samples from children, reported control yet unhealthy samples, and those of patients undergoing antibiotic treatment. The COVID-19 cohort can be found under the accession PRJEB64515 in the European Nucleotide Archive. The Q2PD is deposited on GitHub and can be accessed along with sample accessions and scripts needed to reproduce our results here: [18]. The complete code and additional data required to reproduce our analyses have also been deposited as a GigaDB dataset [43]. DOME-ML (Data, Optimization, Model and Evaluation in Machine Learning) annotations are available in the DOME registry via accession 4cstv1dfjm [44]. An archival copy of the code is available via Software Heritage [45].

Competing Interests

All authors declare no competing interests.

References

1. M'koma AE. Inflammatory bowel disease: an expanding global health problem. *Clin Med Insights Gastroenterol*. 2013;6:CGast.S12731. <https://doi.org/10.4137/CGast.S12731>.
2. Hong Y-R, Huo J, Desai R, et al. Excess costs and economic burden of obesity-related cancers in the United States. *Value Health*. 2019;22(12):1378–86. <https://doi.org/10.1016/j.jval.2019.07.004>.
3. Durack J, Lynch SV. The gut microbiome: relationships with disease and opportunities for therapy. *J Exp Med*. 2019;216(1):20–40. <https://doi.org/10.1084/jem.20180448>.
4. Bajaj JS, Acharya C, Sikaroodi M, et al. Cost-effectiveness of integrating gut microbiota analysis into hospitalisation prediction in cirrhosis. *GastroHep*. 2020;2(2):79–86. <https://doi.org/10.1002/ygh2.390>.
5. DeGruttola AK, Low D, Mizoguchi A, et al. Current understanding of dysbiosis in disease in human and animal models. *Inflamm Bowel Dis*. 2016;22(5):1137–50. <https://doi.org/10.1097/MIB.0000000000000750>.
6. Hrnčir T. Gut microbiota dysbiosis: triggers, consequences, diagnostic and therapeutic options. *Microorganisms*. 2022;10(3):578. <https://doi.org/10.3390/microorganisms10030578>.
7. Martinez JE, Kahana DD, Ghuman S, et al. Unhealthy lifestyle and gut dysbiosis: a better understanding of the effects of poor diet and nicotine on the intestinal microbiome. *Front Endocrinol*. 2021;12:667066. <https://www.frontiersin.org/articles/10.3389/fendo.2021.667066>.
8. Bull MJ, Plummer NT. Part 2: treatments for chronic gastrointestinal disease and gut dysbiosis. *Integr Med Clin J*. 2015;14(1):25–33.
9. Li J, Bates KA, Hoang KL, et al. Experimental temperatures shape host microbiome diversity and composition. *Global Change Biol*. 2023;29:41–56. <https://doi.org/10.1111/gcb.16429>.
10. Gong D, Gong X, Wang L, et al. Involvement of reduced microbial diversity in inflammatory bowel disease. *Gastroenterol Res Pract*. 2016;2016:1. <https://doi.org/10.1155/2016/6951091>.
11. Zouiouich S, Loftfield E, Huybrechts I, et al. Markers of metabolic health and gut microbiome diversity: findings from two population-based cohort studies. *Diabetologia*. 2021;64(8):1749–59. <https://doi.org/10.1007/s00125-021-05464-w>.
12. Gupta VK, Kim M, Bakshi U, et al. A predictive index for health status using species-level gut microbiome profiling. *Nat Commun*. 2020;11(1):Art. No. 1. <https://doi.org/10.1038/s41467-020-18476-8>.
13. Chang D, Gupta VK, Hur B, et al. Gut Microbiome Wellness Index 2 enhances health status prediction from gut microbiome taxonomic profiles. *Nat Commun*. 2024;15(1):7447. <https://doi.org/10.1038/s41467-024-51651-9>.
14. Zhu J, Xie H, Yang Z, et al. Statistical modeling of gut microbiota for personalized health status monitoring. *Microbiome*. 2023;11(1):184. <https://doi.org/10.1186/s40168-023-01614-x>.
15. Berg G, Rybakova D, Fischer D, et al. Microbiome definition re-visited: old concepts and new challenges. *Microbiome*. 2020;8(1):103. <https://doi.org/10.1186/s40168-020-00875-0>.
16. Lee J-Y, Bays DJ, Savage HP, et al. The human gut microbiome in health and disease: time for a new chapter? *Infect Immun*. 2024;92(11):e0030224. <https://doi.org/10.1128/iai.00302-24>.
17. Lloyd-Price J, Arze C, Ananthakrishnan AN, et al. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature*. 2019;569:655–62. <https://doi.org/10.1038/s41586-019-1237-9>.
18. q2-predict-dysbiosis [Online]. <https://github.com/Kizielins/q2-predict-dysbiosis>. Accessed 10 November 2024.
19. Zou Y, Xue W, Luo G, et al. 1,520 reference genomes from cultivated human gut bacteria enable functional microbiome analyses. *Nat Biotechnol*. 2019;37(2):179–85. <https://doi.org/10.1038/s41587-018-0008-8>.
20. Segata N, Izard J, Waldron L, et al. Metagenomic biomarker discovery and explanation. *Genome Biol*. 2011;12(6):R60. <https://doi.org/10.1186/gb-2011-12-6-r60>.
21. Mirsepasi-Lauridsen HC, Vrankx K, Engberg J, et al. Disease-specific enteric microbiome dysbiosis in inflammatory bowel disease. *Front Med*. 2018;5:304. <https://www.frontiersin.org/articles/10.3389/fmed.2018.00304>.
22. Mosca A, Leclerc M, Hugot JP. Gut microbiota diversity and human diseases: should we reintroduce key predators in our ecosystem? *Front Microbiol*. 2016;7:455. <https://www.frontiersin.org/articles/10.3389/fmicb.2016.00455>.
23. Piliszek R, Mnich K, Migacz S, et al. MDFS: multiDimensional feature selection in R. *R J*. 2019;11(1):198–210. <https://doi.org/10.32614/RJ-2019-019>.
24. Mnich K, Rudnicki WR. All-relevant feature selection using multidimensional filters with exhaustive search. *Inform Sci*. 2020;524:277–97. <https://doi.org/10.1016/j.ins.2020.03.024>.
25. Friedman J, Alm EJ. Inferring correlation networks from genomic survey data. *PLoS Comput Biol*. 2012;8(9):e1002687. <https://doi.org/10.1371/journal.pcbi.1002687>.
26. Breiman L. Random forests. *Machine Learning*. 2001;45(1):5–32. <https://doi.org/10.1023/A:1010933404324>.

27. Pasolli E, Schiffer L, Manghi P, et al. Accessible, curated metagenomic data through ExperimentHub. *Nat Methods*. 2017;14(11):1023–24. <https://doi.org/10.1038/nmeth.4468>.
28. Kursa MB, Rudnicki WR. Feature selection with the Boruta package. *J Stat Softw*. 2010;36:1–13. <https://doi.org/10.18637/jss.v036.i11>.
29. Zhu C, Miller M, Marpa S, et al. Functional sequencing read annotation for high precision microbiome analysis. *Nucleic Acids Res*. 2018;46(4):e23. <https://doi.org/10.1093/nar/gkx1209>.
30. Maranga M, Szczerbiak P, Bezshapkin V, et al. Comprehensive functional annotation of metagenomes and microbial genomes using a deep learning-based method. *mSystems*. 2023;8(2):e01178–22. <https://doi.org/10.1128/msystems.01178-22>.
31. Koehler Leman J, Szczerbiak P, Renfrew PD, et al. Sequence-structure-function relationships in the microbial protein universe. *Nat Commun*. 2023;14(1):Art. No. 1. <https://doi.org/10.1038/s41467-023-37896-w>.
32. Szydlowski LM, Bulbul AA, Simpson AC, et al. Adaptation to space conditions of novel bacterial species isolated from the International Space Station revealed by functional gene annotations and comparative genome analysis. *Microbiome*. 2024;12:190. <https://doi.org/10.1186/s40168-024-01916-8>.
33. Kańduła MM, Aldoshin AD, Singh S, et al. ViLoN—a multi-layer network approach to data integration demonstrated for patient stratification. *Nucleic Acids Res*. 2023;51(1):e6. <https://doi.org/10.1093/nar/gkac988>.
34. Krueger F. *Trim Galore*. 2023. Perl. <https://github.com/FelixKrueger/TrimGalore>. Accessed 4 December 2023.
35. Blanco-Míguez A, Beghini F, Cumbo F, et al. Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4. *Nat Biotechnol*. 2023;41(11):1633–44. <https://doi.org/10.1038/s41587-023-01688-w>.
36. Beghini F, Mciver LJ, Blanco-Míguez A, et al. Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. *eLife*. 2021;10:e65088. <https://doi.org/10.7554/eLife.65088>.
37. Bolyen E, Rideout JR, Dillon MR, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol*. 2019;37(8):852–57. <https://doi.org/10.1038/s41587-019-0209-9>.
38. SCNIC. <https://github.com/lozuponelab/q2-SCNIC>. Accessed 10 November 2024.
39. Leo Breiman JF. *Classification and Regression Trees*. Boca Raton, FL: Chapman and Hall/CRC, 1984.
40. Fernandez-Delgado M, Cernadas E, Barro S, et al. Do we need hundreds of classifiers to solve real world classification problems? *J Mach Learn Res*. 2014;15(1):3133–81. <https://dl.acm.org/doi/10.5555/2627435.2697065>.
41. q2-health-index. <https://github.com/Kizielins/q2-predict-dysbiosis>. Accessed 10 November 2024.
42. Li H. *lh3/seqtk*. 2023. <https://github.com/lh3/seqtk>. Accessed 4 December 2023.
43. Zielińska K, Udekwu KI, Rudnicki W, et al. Supporting data for “Healthy Microbiome—Moving toward Functional Interpretation.” *GigaScience Database*. 2025. <https://doi.org/10.5524/102656>.
44. Zielińska K, Udekwu KI, Rudnicki W, et al. Healthy microbiome—moving towards functional interpretation [DOME-ML Annotations]. *DOME-ML Registry*. 2025. <https://registry.dome-ml.org/revview/4cstv1dfjm>. Accessed 20 January 2025.
45. Zielińska K, Udekwu KI, Rudnicki W, et al. Healthy microbiome—moving towards functional interpretation (Version 1) [Computer software]. *Software Heritage*. 2025. <https://archive.softwareheritage.org/wh:1:snpc8fff6424c577c1495cd04c81a21657540563131>. Accessed 20 February 2025.