

RESEARCH ARTICLE

The importance of age in compositional and functional profiling of the human intestinal microbiome

Elio L. Herzog^{1,2}, Melania Wäfler¹, Irene Keller^{3,4}, Sebastian Wolf^{1,3}, Martin S. Zinkernagel^{1,3}, Denise C. Zysset-Burri^{1,3*}

1 Department of Ophthalmology, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland, **2** Graduate School for Cellular and Biomedical Sciences, University of Bern, Bern, Switzerland, **3** Department for BioMedical Research, University of Bern, Bern, Switzerland, **4** Interfaculty Bioinformatics Unit and Swiss Institute of Bioinformatics, University of Bern, Bern, Switzerland

✉ These authors contributed equally to this work.

* denise.zysset@insel.ch



OPEN ACCESS

Citation: Herzog EL, Wäfler M, Keller I, Wolf S, Zinkernagel MS, Zysset-Burri DC (2021) The importance of age in compositional and functional profiling of the human intestinal microbiome. PLOS ONE 16(10): e0258505. <https://doi.org/10.1371/journal.pone.0258505>

Editor: Brenda A. Wilson, University of Illinois Urbana-Champaign, UNITED STATES

Received: April 27, 2021

Accepted: September 24, 2021

Published: October 18, 2021

Copyright: © 2021 Herzog et al. 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 use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript, its [Supporting Information](#) files, and available in the European Nucleotide Archive under accession numbers PRJEB13835, PRJEB24557, PRJEB35615, PRJEB38145, PRJEB38997 and PRJEB44233.

Funding: This work was supported by OPOS foundation, St. Gallen, Switzerland, <https://www.fundraiso.ch/sponsor/stiftung-zur-foerderung-der-ostschweizerischen-pleoptik-und-orthoptikschule-opos-in-st-gallen>, PI: D.C.Z.-B., and Fondation

Abstract

The intestinal microbiome plays a central role in human health and disease. While its composition is relatively stable throughout adulthood, the microbial balance starts to decrease in later life stages. Thus, in order to maintain a good quality of life, including the prevention of age-associated diseases in the elderly, it is important to understand the dynamics of the intestinal microbiome. In this study, stool samples of 278 participants were sequenced by whole metagenome shotgun sequencing and their taxonomic and functional profiles characterized. The two age groups, *below65* and *above65*, could be separated based on taxonomic and associated functional features using Multivariate Association of Linear Models. In a second approach, through machine learning, biomarkers connecting the intestinal microbiome with age were identified. These results reflect the importance to select age-matched study groups for unbiased metagenomic data analysis and the possibility to generate robust data by applying independent algorithms for data analysis. Furthermore, since the intestinal microbiome can be modulated by antibiotics and probiotics, the data of this study may have implications on preventive strategies of age-associated degradation processes and diseases by microbiome-altering interventions.

Introduction

Genetic diversity between humans does not only arise from allele frequency differences of shared human genes, but also from the vast number of genetic and metabolic diversity in intestinal microbial communities. The human intestinal microbiome is a complex system consisting of trillions of microorganisms that contribute to numerous functions of the host. Fermentation of indigestible food components, stimulation and regulation of the immune system, strengthening of the intestinal barrier and protection against pathogens [1] are some of the key functions of the intestinal microbiome. Despite its crucial role in human health, the

Bertarelli Catalyst Fund, EPFL (Ecole Polytechnique Fédérale de Lausanne), Lausanne, Switzerland (CF10000044 – EPFL SCRO237812), <https://www.epfl.ch/research/services/fund-research/funding-opportunities/research-funding/catalyst-fund/>, PI: D.C.Z.-B. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

composition of the intestinal microbiome is not uniform between individuals and populations [2]. Although there are significant differences in relative abundances of bacteria between individuals, the phyla *Bacteroidetes*, *Firmicutes* and *Proteobacteria* seem to dominate the composition in almost all individuals [3]. However, despite being able to remain stable over decades in an individual [4], the composition of the human intestinal microbiome is influenced by several factors, including the genetic background, gut architecture, immune system, body mass index (BMI), diet, life style, antibiotics intake, disease and age [5].

With increasing age, physiological functions of human organs start to gradually decrease [6], making them prone to infections and diseases and leading to a higher mortality risk in the elderly population [7, 8]. The gastrointestinal tract is also vulnerable to this aging process. Thus, understanding its age-related dynamics may be crucial for disease prevention in the elderly. The influence of age on the microbial composition in the gut has been investigated in many studies for over a decade [9–13]. The most noticeable feature in the microbiota of elderly individuals is an altered ratio of *Firmicutes* to *Bacteroidetes*, with an increased proportion of *Bacteroidetes* in the elderly [10]. This ratio has been shown to be of significant relevance in several disease states [14] and seems to be not only an effect of the current age, but can, in turn, have an impact on the ageing process itself [15, 16].

In this study, we aimed to investigate age-associated changes in the intestinal microbial composition and in microbial functional profiles. Stool samples of 278 participants were sequenced and analyzed using two independent approaches to identify associations between microbial abundances as well as functional profiles and age. In contrast to many previous studies, whole metagenome shotgun sequencing instead of 16S ribosomal RNA sequencing was applied, allowing the identification of archaea, viruses and eukaryotes in addition to bacteria. Since the stability of the intestinal microbiome diminishes between an age of 63 and 76 years [12, 13], we set a threshold of 65 years to divide the cohort into two sex-matched age groups *below65* and *above65*.

Materials and methods

Study design and recruitment

Participants (n = 278) were recruited from the Department of Ophthalmology at the University Hospital Bern (Inselspital), Switzerland. The study follows the ethical principles for medical research found in the Declaration of Helsinki and was approved by the Ethics Committee of the Canton of Bern (Clinical-Trials.gov: NCT02438111). After receiving oral and written information, all participants signed the informed consent prior to participation. We tested for differences between study groups in a range of demographic values using either Welch's t-test (for age and BMI) or Fisher's exact test (for sex and smoking; Table 1). Exclusion criteria were chronic inflammatory and gastrointestinal diseases (including previous surgery in the gastrointestinal tract) as well as systemic antibiotic treatment within the last three months. All participants were Caucasian and were 18 years of age or older.

Sample collection, data sequencing and quality control

Chilled stool samples were collected and delivered in an aerobic environment and brought to the laboratory within 16 hours after defecation. Upon arrival, they were immediately frozen at -20°C. Following the manufacturer's protocol, the PSP® Spin Stool DNA Plus kit (Stratec Biomedical AG, Beringen, Switzerland) with an integrated RNA digestion step using 100 mg/ml RNase A (Qiagen, Hombrechtikon, Switzerland) was used to isolate metagenomic DNA from up to 200 mg stool sample. Whole metagenome shotgun sequencing was performed at BGI Europe (Copenhagen N, Denmark) and the Next Generation Sequencing Platform of the

Table 1. Characterization of the study cohort.

Feature	above 65	below 65	p-value
N	145	133	
Sex (% male)	64 (44.1)	67 (50.4)	0.357
Age [years] (mean (SD), min, max)	77.25 (6.32), 65.07, 92.94	49.69 (12.05), 19.55, 64.71	<0.001
BMI [kg/m ²] (mean (SD))	26.10 (4.54)	24.80 (4.44)	0.016
Smoker (%)			0.058
No	85 (58.6)	74 (55.6)	
Previous	53 (36.6)	42 (31.6)	
Yes	7 (4.8)	17 (12.8)	

BMI = Body Mass Index [kg/m²]. Differences in BMI and age were calculated using Welch's t-test, differences in sex and smoking through Fisher's exact test.

<https://doi.org/10.1371/journal.pone.0258505.t001>

University of Bern, Switzerland. For library preparation, the TruSeq DNA PCR-Free Library Preparation kit was used. Cluster generation and sequencing were done following standard pipelines of Illumina HiSeq 3000 platforms, resulting in 150bp paired-end reads. Quality filtering was performed using Trimmomatic v.0.32 to remove adapter sequences and reads shorter than 70bp and to trim low-quality bases from both ends [17]. Resulting reads were mapped against hg19 human reference genome to identify sequences of human origin using Bowtie2 (v.2.2.4) [18]. Reads of human origin were excluded, resulting in non-human, high-quality reads for further analysis.

Microbial and functional profiling of the intestinal microbiomes

The Metagenomic Phylogenetic Analysis tool (MetaPhlAn2, v.2.0–2.6.0) [19] and the marker database (v.20) using default settings were used to perform metagenomic profiling by mapping non-human high-quality reads to a set of clade-specific markers. Alignment was performed using Bowtie2 (v.2.2.4) followed by normalization of the total number of reads in each clade divided by the nucleotide length of its marker, resulting in the relative abundance of each taxonomic unit.

To detect the metabolic potential of the gut microbiome, the HMP (Human Microbiome Project [20]) Unified Metabolic Analysis Network (HUMAN2, v.0.2.1 – v.0.11.0) [21] was applied for each sample separately with default settings based on the taxonomic profiles from MetaPhlAn2. Mapping reads to ChocoPhlAn, a functionally annotated pan-genome database, was performed with the help of Bowtie2 (v.2.2.4). To identify unmapped reads, Diamond (v.0.8.37) [22] in combination with the universal protein reference database UniRef90 [23] was used. The assignment of the resulting organism-specific gene hits to pathways was done through maximum parsimony using MinPath (v.1.2) [24]. Using this information, HUMAN2 returned a list of genes and pathways and their relative abundances.

A Principle Component Analysis (PCA) between groups was computed and visualized using the function *prcomp* (data, center = T, scale = T) [25] and the library *factoextra* in the R software (version 3.6.0) [26]. PCA was performed for microbial abundances (Fig 3A) and pathway abundances (Fig 3B). The p-value for separation was assessed by Permutation Multi-variate Analysis of Variance (PERMANOVA) with 10'000 iterations using the R package *vegan51* [27]. To find age-related taxonomic and functional features, Multivariate Association of Linear Models (MaAsLin) in the R package *Maaslin* (v.0.0.5) was used with default settings [28]. Moreover, associations of biological variables including sex and BMI with microbial and functional abundances were analyzed with MaAsLin. Differences were considered to be significant

if the p-value < 0.05 and q-value < 0.2. MetaPhlAn2 abundance files were normalized by summing all values across all taxonomic levels for each participant followed by dividing each value by this sum. Selection was performed by maintaining only the most specific taxa if abundances of one clade of several taxonomic levels were identical for all participants.

Group separation by machine learning

A machine learning approach was used to investigate a potential separation of the study groups *below65* and *above65* by their microbial and functional profiles and to find the main contributors of such a separation. Model selection for the best performing machine learning algorithm of the dataset was done with the R libraries *mlbench* (v.2.1) [29] and *caret* (v.6.0–84) [30], testing four common classification algorithms: CART (Classification and Regression Tree), SVM (Supported Vector Machines), RF (Random Forest) and KNN (K nearest neighbor). The best fitting model, R package *randomForest* (v.4.6–14), was consequentially used [31].

Parameter tuning of random forest was performed for *mtry* and *nTree* using random and grid search as well as build in tools. The algorithm's performance on the whole dataset using 10 fold cross validation was calculated using the package *caret* (v.6.0–84) [30]. Cross-validation was repeated 10 times. Random forests performance was evaluated by fitting it on the training set using the fitted model and the function *predict* of the random Forest package. Receiver Operating Characteristic (ROC) curves were calculated using the R package *ROCR* (v.1.0–7) [32]. Shrinkage Discriminant Analysis (SDA) based on Correlation Adjusted T-scores (CAT-scores) was performed using the R package *sda* (v.1.3.7) [33]. A shrinkage CAT score between the mean values of the groups was computed for each predictor variable. The ranking for each feature was determined by a summary score (the weighted sum of squared CAT scores across classes) using microbial and pathway abundances as input. Based on Gene Ontology (GO) terms, functional features with top CAT scores were clustered using REVIGO [34].

Results

Taxonomic and functional characterization of the intestinal microbiota

To find associations of age with functional and compositional alterations in the intestinal microbiome, the gut metagenomes of 278 study participants were sequenced. The cohort consisted of 145 participants aged equal or above 65 years (*above65*) and 133 participants aged below 65 years (*below65*) (Table 1). In total, 7.3 trillion 151 bp paired-end reads with an average of 26 ± 10.9 (SD) million reads per sample were generated. After trimming and filtering, 23 ± 9.9 (SD) million non-human high-quality reads per sample remained for further processing. Overall, 99.43% of the reads mapped to the bacterial kingdom (99.88% in participants *below65*, 99.01% in participants *above65*). *Bacteroidetes* and *Firmicutes*, followed by *Proteobacteria* and *Actinobacteria* were found to be the most abundant phyla (Fig 1B). Consistent with previous studies, *Bacteroidia* and *Clostridia* were the most abundant classes in the cohort [35]. The dominating genera were *Bacteroides*, *Alistipes* followed by *Subdoligranulum* and *Faecalibacterium*. An unclassified *Subdoligranulum* species, *Faecalibacterium prausnitzii*, *Alistipes putredinis*, *Prevotella copri* and *Bacteroides uniformis* were found to be the five most abundant species in the cohort (S1 Table). To describe the metabolic functions of these identified taxa, HUMAnN2 was applied on each sample separately, resulting in 793 assigned pathways.

Classification of the microbiota into enterotypes

In accordance with a previous study of Arumagam *et al.* [35], the intestinal microbiomes of our cohort could be divided into three enterotypes of distinct microbial composition. Out of

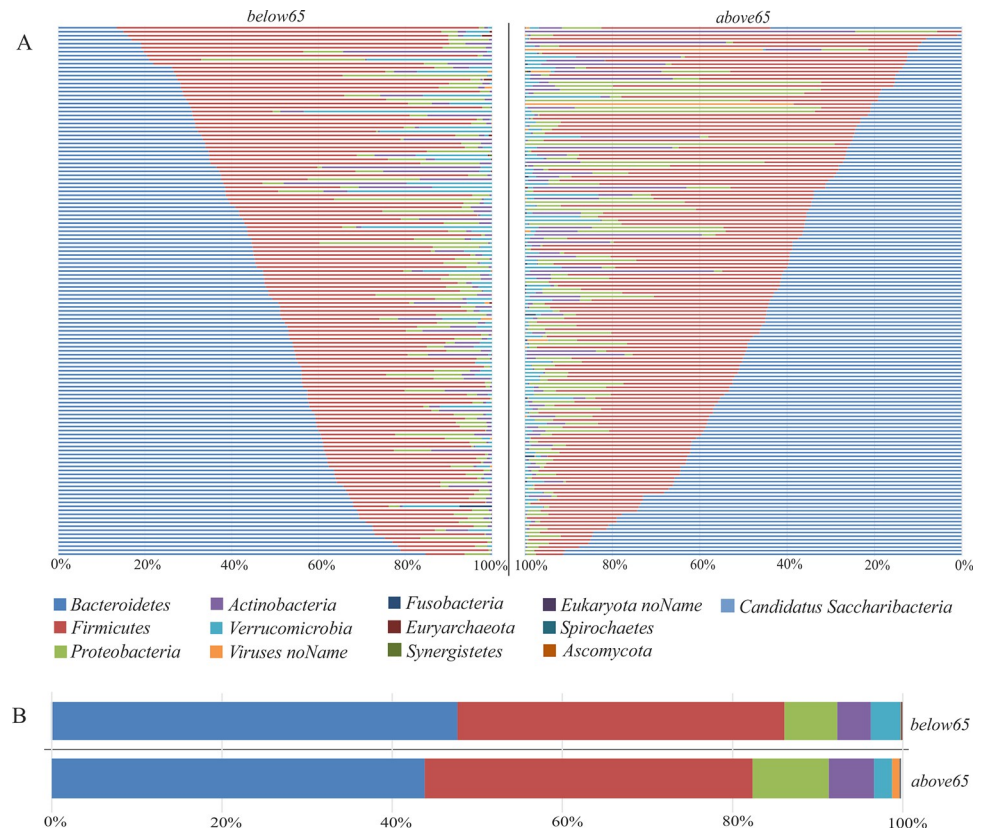


Fig 1. Taxonomic characterization of the intestinal microbiome. Relative abundances of microbiota at phylum level for each study subject (A) and averaged for study groups (B). *Above65* (patients aged 65 years and above; n = 145), *below65* (patients below 65 years of age; n = 133).

<https://doi.org/10.1371/journal.pone.0258505.g001>

the computed Jensen-Shannon distance of the genus abundances, clustering was done with Partitioning Around Medoids (PAM). The Calinski-Harabasz (CH) Index was used to determine that the optimal cluster number, i.e. returning the most robust partition of the dataset, was three clusters. Combining the results of a PCA and clustering through Between Class Analysis (BCA) resulted in graphical interpretation of the data in Fig 2A. In terms of abundance *Bacteroides*, *Prevotella* and *Subdoligranulum* were found to be the dominating genera in clusters 1, 2 and 3, respectively (Fig 2B). Applying Fisher's exact test, subjects of group *above65* were over-represented in enterotype 2 ($p = 0.0012$) and under-represented in enterotype 3 ($p = 0.0036$), supposing an age-dependency of the proposed enterotypes.

Age-dependent microbial and functional composition of the intestinal microbiota

A PCA with age as grouping variable showed that differences in microbial species abundance as well as in pathway abundance separated the two age groups *above65* and *below65* with PERMANOVA confirming a significant p-value of 0.0004 and 0.0006, respectively (10'000 iterations; Fig 3A and 3B). To identify features that are different in relative abundance between the groups, MaAsLin was applied on the taxonomically and functionally profiled metagenomes. Out of the 20 identified taxa with age-dependence, 15 had a higher relative abundance in the *below65* group and five in the *above65* group (Fig 3C). Moreover, while the intestinal microbiomes of subjects of the *below65* group were enriched in genes of 248 pathways, microbiomes

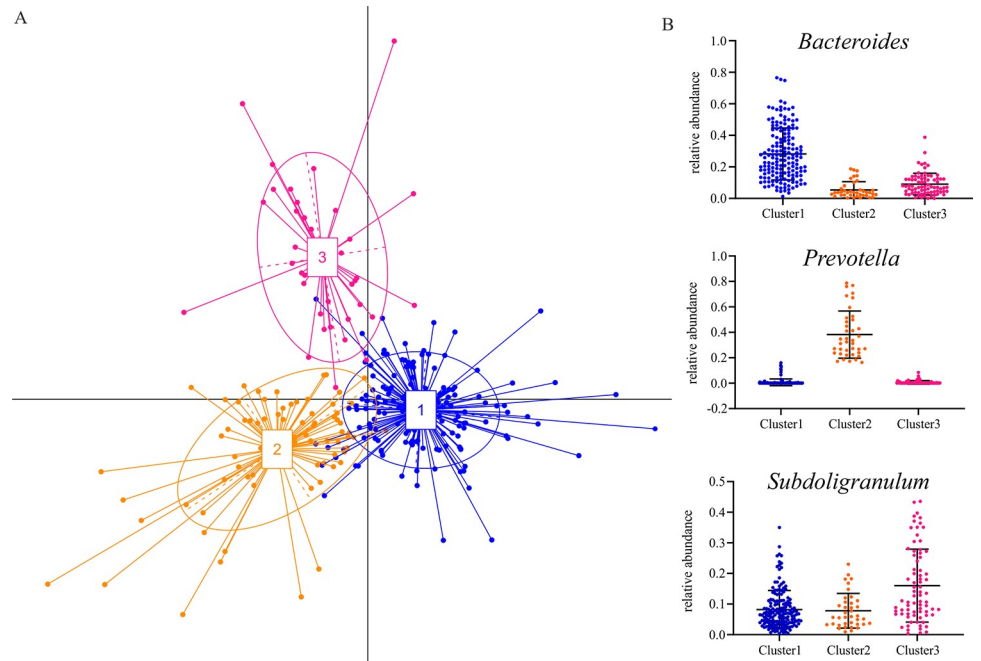


Fig 2. Intestinal microbial enterotypes. (A) Based on the abundance of microbial genera, three enterotypes were identified in the cohort using Between Class Analysis that visualizes results from Principal Component Analysis and clustering. (B) The relative abundances of the proposed drivers of these three enterotypes, the genera *Bacteroides*, *Prevotella* and *Subdoligranulum*, are shown for each subject.

<https://doi.org/10.1371/journal.pone.0258505.g002>

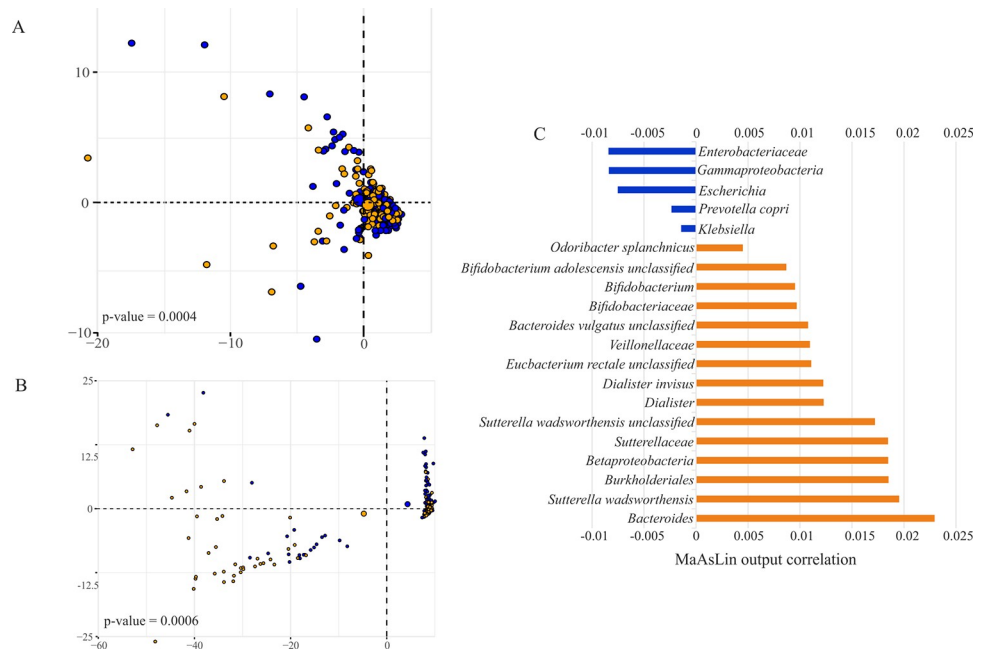


Fig 3. Distinct microbial and functional composition between age groups. Principal component analysis of (A) microbial and (B) pathway abundances separated the two age groups *above65* and *below65* (PERMANOVA, 10'000 iterations). Blue color represents *above65* (patients aged 65 years and above; n = 145), orange *below65* (patients below 65 years of age; n = 133). (C) Correlation between taxonomic features and age (MaAsLin, $q < 0.2$). Positive correlations (orange) imply higher abundance in *below65*, whereas negative correlations (blue) imply higher abundance in *above65*.

<https://doi.org/10.1371/journal.pone.0258505.g003>

of subjects of the *above65* group were enriched in genes of 57 pathways (S2 Table). A boosting step in the MaAsLin algorithm ensures that only metadata that are associated with the given metagenomic feature are included in the model, implying that all associations detected by the modeling approach have been corrected for all other confounding factors. However, the effect of other biological variables including sex and BMI on the microbiome of this cohort was also investigated, showing that only a subspecies of *Bacteroidales bacterium ph8* was associated with sex with higher abundances in females compared to males and that the BMI positively correlated with the order *Selenomonadales* and negatively correlated with the family *Ruminococcaceae*. There was no association found between both, sex and BMI, and the functional profiles of the metagenomes in the cohort.

Allocation of subjects to age groups based on the intestinal microbiome

To further illustrate the age-dependency of the intestinal microbiome, machine learning approaches were applied to identify potential biomarkers for the age groups. Model selection of several common machine learning algorithms for classification revealed Random Forest as most suitable for the data set since it showed significantly better performance for both, microbial and pathway abundances, compared to the other algorithms tested (S1 Fig). Furthermore, to increase the accuracy of Random Forest, hyperparameters were tuned, suggesting setting the number of variables used in each split (*mtry*) to 11 and the number of generated trees in the forest (*nTree*) to 2000 for taxonomic features and setting *mtry* to 26 and *nTree* to 2000 for functional features, respectively. After tuning, the model was trained and evaluated by 10 fold cross-validation (Table 2). While accuracies between 0.72 and 0.86 mean that between 72 and 86% of the features are classified correctly by the Random Forest approach, kappa between 0.44 and 0.71 indicates moderate to substantial agreement according to Landis and Koch [36]. To further assess the classifier's performance, ROC curves were computed (Fig 4). With an AUC of 0.91 for microbial abundances (Fig 4A) and 0.77 for pathway abundances (Fig 4B), the discrimination capacities of the models to distinguish between age groups is good. Finally, to identify those taxonomic and functional features contributing most to group separation, shrinkage discriminant analysis based on CAT-scores was applied (Fig 5). According to *mtry* in Random Forest, 11 bacterial groups (Fig 5A) and 26 pathways (Fig 5B) with the highest CAT-scores were considered. GO term based clustering revealed that age-dependent functional features of the microbiome are mainly involved in biosynthetic processes of heme, sphingolipid, unsaturated fatty acids and nicotine catabolic process (Fig 6).

Discussion

In this study, the effects of age on the intestinal microbiome and its functional profile were investigated. The intestinal microbiome is relatively stable during adulthood [4], but several studies reported aberrations in older individuals [11, 37]. In accordance to previous studies

Table 2. Performance of random forest on microbial and pathway abundances.

	Microbial abundances	Pathway abundances
Sample size	278	277
Accuracy	0.755	0.724
Kappa	0.507	0.442

The model was evaluated by ten times repeated 10-fold cross-validation. Accuracy shows the number of correctly classified features out of all features, while kappa depicts the accuracy normalized at the baseline of random chance on the dataset.

<https://doi.org/10.1371/journal.pone.0258505.t002>

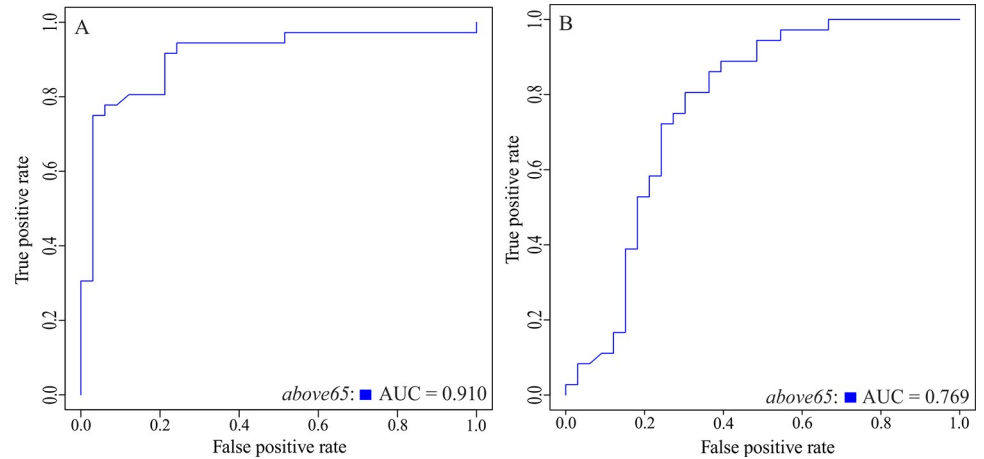


Fig 4. ROC curves for the random forest classifier for microbial abundances and pathway abundances. ROC curves visualizing the Random Forest classifier for (A) microbial abundances and (B) pathway abundances. The group *above65* is represented in blue (patients aged 65 years and above; n = 145), the curve for *below65* was omitted as its information is redundant. AUC, area under the curve; ROC, receiver operating characteristic.

<https://doi.org/10.1371/journal.pone.0258505.g004>

[35, 38], the microbiomes in this cohort were dominated by the phyla *Bacteroidetes* and *Firmicutes* (Fig 1). However, differences in relative abundances of microbes between the age groups *below65* and *above65* compared in this study, have been identified by two independent approaches, by MaAsLin (Fig 3C) and by machine learning algorithms using Random Forest

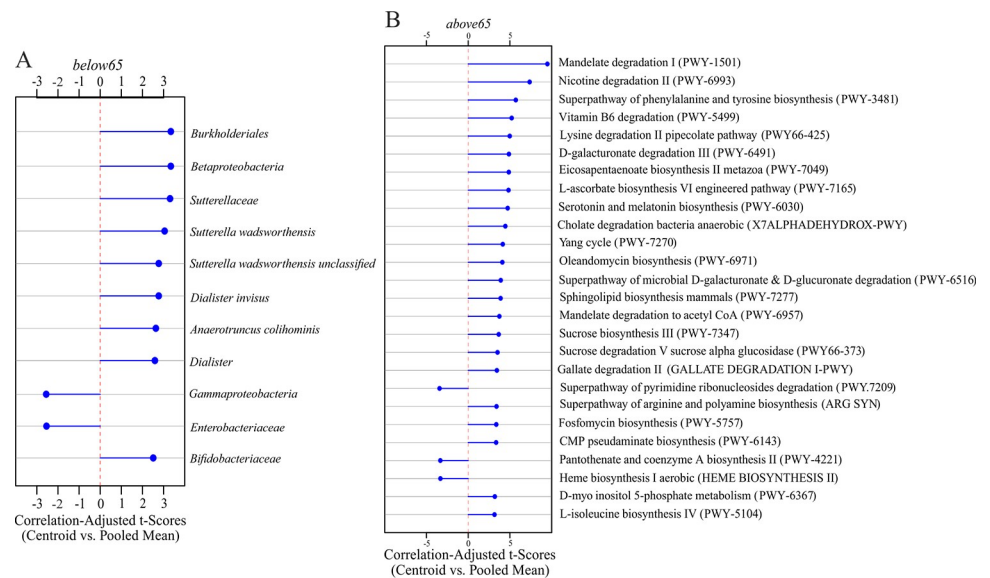


Fig 5. Top 11 bacteria and top 26 pathways according to CAT-scores. The (A) top ranked 11 bacteria and (B) top ranked 26 pathways according to CAT-scores. The length and direction of the blue bars indicate the influence of a given biomarker on the discriminative power of the model. A: The order *Burkholderiales* within the class *Betaproteobacteria* have the highest potential for separation of the age groups with a positive CAT-score indicating an over-representation in the *below65* group. While both, the family *Enterobacteriaceae* within the class *Gammaproteobacteria* have a negative CAT-score indicating an over-representation in the *above65* group. B: Most of the pathways shown have a higher relative abundance in the elderly population with mandelate degradation (PWY-1501) contributing most to group separation. *above65*: patients aged 65 years and above (n = 145), *below65*: patients below 65 years of age (n = 133). CAT-scores, correlation adjusted T-scores.

<https://doi.org/10.1371/journal.pone.0258505.g005>

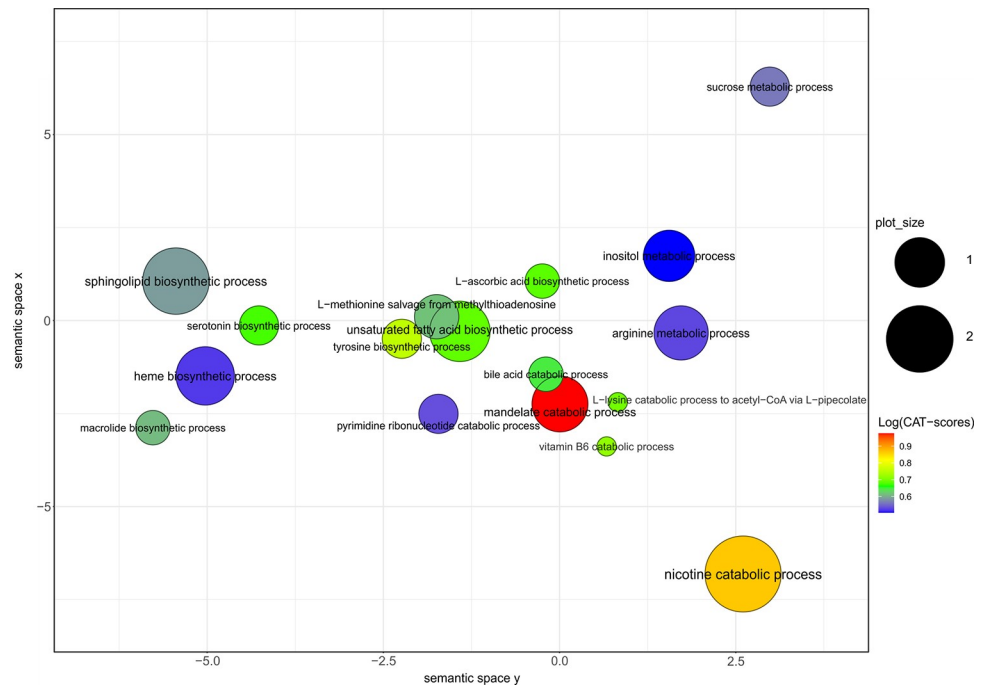


Fig 6. Scatterplot of the top 26 pathways based on GO terms. Cluster representatives (i.e. GO terms remaining after redundancy reduction by REVIGO) are shown. Distances between bubbles represent the semantic similarities between the GO terms, bubble position is determined by application of multidimensional scaling to a matrix of the GO terms' semantic similarities (the lower the distance, the more similar the terms). The axes values have no intrinsic meaning. Bubble size indicates the frequency of the GO term in the underlying GO database. Logarithmized CAT-scores are visualized using a color gradient from red to blue. GO, gene ontology; CAT-scores, correlation adjusted T-scores.

<https://doi.org/10.1371/journal.pone.0258505.g006>

(Fig 5A). Using PCA, the group separation may in part be due to outliers since most of the data clusters together (Fig 3A and 3B). However, both approaches showed that the class *Beta-proteobacteria* as well as its order *Burkholderiales* and the family *Sutterellaceae* including the species *Suturella wadsworthensis* had a higher relative abundance in the *below65* age group, whereas the class *Gammaproteobacteria* and its family *Enterobacteriaceae* had a higher relative abundance in the *above65* age group compared to the respective other group.

Although being in relatively low abundance compared to *Bacteroidetes* and *Firmicutes*, alterations in the phylum *Proteobacteria* may have a considerable effect on human health since an elevated prevalence of *Proteobacteria* has been proposed as a diagnostic marker for an unstable intestinal microbial community called dysbiosis and for risk of disease [38–40]. Moreover, the family *Bifidobacteriaceae* and its genus *Bifidobacterium* including the species *Bifidobacterium adolescentis* were of higher relative abundance in the *below65* compared to the *above 65* age group. Since it has been proposed in many studies that *Bifidobacterium* can be used as probiotic to alleviate various disease by modulating the intestinal microbial composition [41], reduced *Bifidobacteriaceae* may be used as marker for dysbiosis and disease progression in the elderly. Furthermore, increasing proportions of *Enterobacteriaceae* as observed in the *above65* group, including *Klebsiella* spp., *Enterobacter aerogenes* and *Escherichia coli*, were also observed in patients suffering from atherosclerotic cardiovascular disease [42]. Bacteria of the genus *Klebsiella* have been observed in higher abundances in patients with hypertension and pre-hypertension [43]. Thus, the family *Enterobacteriaceae* and especially its genus *Klebsiella* may be a marker for disease in the elderly.

Concerning taxonomic features, the two approaches applied for age group separation in this study, resulted in similar results with some exceptions at low taxonomic levels. On the level of functional features, machine learning algorithms assisted in reducing the data set into pathways with the highest discriminative potential in the prediction model of the cohort. Whereas MaAsLin resulted in 305 significantly different pathways between *below65* and *above65* (S2 Table), the use of Random Forest approach allowed to reduce this list to the 26 pathways contributing most to group separation (Fig 5B). This shows that machine learning is a powerful tool to find key differences in the data set. However, there are also disadvantages: Unlike to simple p-value test, relevance scores used to highlight multivariate interacting effects in machine learning approaches are usually difficult to interpret [44]. Moreover, through bootstrapping of the dataset some samples may be lost due to random sampling, leading to possible neglect of crucial data such as outliers. Using machine learning to develop classifiers for disease detection has the advantage in its non-invasive nature, but it is crucial to use attributes resulting in classifiers with reasonable predictive value for disease instead of confounding variables [45]. Therefore, combining different approaches, termed hybrid machine learning, may result in more stable and better predicting algorithms to detect potential biomarkers [46]. In this study, only one machine learning approach (Random Forest that showed the best performance in the prediction model, S1 Fig) was applied, but a second independent algorithm based on linear models (MaAsLin) was used for data analysis. This may mutually exclude the drawbacks of the two approaches and generate more robust results.

In this study, the highest discriminative power in the proposed model was attributed to 26 pathways that are mainly involved in biosynthetic processes of heme, sphingolipid, unsaturated fatty acids and nicotine catabolic process (Fig 6). Many studies have shown associations between the Firmicutes to Bacteroidetes ratio and several diseases such as obesity [47] and also including age-dependent diseases such as age-related macular degeneration [48]. Since the order Selenomonadales is part of the phylum Firmicutes, the positive correlation found between Selenomonadales and the BMI in this cohort may point to these associations. Moreover, the fatty acid profile is linked to various metabolic disorder including obesity [49]. Thus, there may be an age-associated connection between the Firmicutes to Bacteroidetes ratio, fatty acid synthesis and several diseases such as obesity. Molano et al. showed age-dependent changes in the sphingolipid composition of immune cells, resulting in immune dysregulation [50]. Thus, the age-dependent sphingolipid biosynthesis by gut microbes found in this study may be linked to diminished functions of the immune system and associated diseases in the elderly. Moreover, in agreement with our data, it has been shown in a murine model of Western diet in the USA that the microbiome affects both, the plasma fatty acids and the liver sphingolipids [51]. Since the heme metabolism may be altered in age-related diseases, probably involving oxidative damage that is triggered by free heme, and since the biosynthesis of heme requires Vitamin B6 [52], the age-dependent biosynthesis of both, heme and Vitamin B6, found in this study may be a trigger for age-related diseases. Previous studies have shown that nicotine exposure alters the intestinal microbiome secondary to diet [53] and in our cohort, we identified an age-associated nicotine catabolic process, supposing an altered degradation of nicotine in the elderly in association with the taxonomic composition of the microbiome.

Conclusions

This study revealed taxonomic and functional features of the intestinal microbiome associated with age and with a potential link to age-associated diseases in humans. Therefore, these results may have important implications on preventive strategies for degenerative processes occurred in the elderly by using microbiome-altering interventions. Given the significant differences

found between the age groups in this study by two independent approaches, advising to use age-matched groups for unbiased metagenomic data analysis in further studies and to consider the drawbacks of the algorithms used for data analysis, thus, probably applying a second independent approach to generate robust results.

Supporting information

S1 Table. Taxonomic characterization of the intestinal microbiome by MetaPhlan2. The most abundant phyla are highlighted in red, the most abundant classes in green, the dominating genera in purple and the most abundant species in yellow.

(XLSX)

S2 Table. Distinct functional composition between age groups. Correlation between functional features and age (MaAsLin, $q < 0.2$). Positive correlations (orange) imply higher abundance in *below65*, whereas negative correlations (blue) imply higher abundance in *above65*.

(XLSX)

S1 Fig. Model selection for machine learning algorithm based on microbial abundances.

RF, Random Forest; SVM, support vector machine; CART, Classification and Regression Trees; KNN, K-Nearest Neighbor.

(TIF)

Acknowledgments

We thank Prof. Dr. Raphael Sznitman from ARTROG Center for Biomedical Engineering Research of the University of Bern for helpful advice in machine learning approaches.

Author Contributions

Conceptualization: Martin S. Zinkernagel, Denise C. Zysset-Burri.

Formal analysis: Melania Wäfler, Irene Keller, Denise C. Zysset-Burri.

Funding acquisition: Martin S. Zinkernagel, Denise C. Zysset-Burri.

Methodology: Melania Wäfler, Irene Keller.

Supervision: Sebastian Wolf, Martin S. Zinkernagel, Denise C. Zysset-Burri.

Writing – original draft: Elio L. Herzog, Melania Wäfler.

Writing – review & editing: Irene Keller, Sebastian Wolf, Martin S. Zinkernagel, Denise C. Zysset-Burri.

References

1. Heintz-Buschart A. and Wilmes P., Human Gut Microbiome: Function Matters. *Trends Microbiol*, 2018. 26(7): p. 563–574. <https://doi.org/10.1016/j.tim.2017.11.002> PMID: 29173869
2. Nayfach S., et al., New insights from uncultivated genomes of the global human gut microbiome. *Nature*, 2019. 568(7753): p. 505–510. <https://doi.org/10.1038/s41586-019-1058-x> PMID: 30867587
3. Lozupone C.A., et al., Diversity, stability and resilience of the human gut microbiota. *Nature*, 2012. 489(7415): p. 220–30. <https://doi.org/10.1038/nature11550> PMID: 22972295
4. Faith J.J., et al., The long-term stability of the human gut microbiota. *Science*, 2013. 341(6141): p. 1237439. <https://doi.org/10.1126/science.1237439> PMID: 23828941
5. Hall A.B., Tolonen A.C., and Xavier R.J., Human genetic variation and the gut microbiome in disease. *Nat Rev Genet*, 2017. 18(11): p. 690–699. <https://doi.org/10.1038/nrg.2017.63> PMID: 28824167

6. Franceschi C., et al., The extreme longevity: the state of the art in Italy. *Exp Gerontol*, 2008. 43(2): p. 45–52. <https://doi.org/10.1016/j.exger.2007.06.006> PMID: 17703905
7. Troen B.R., The biology of aging. *Mt Sinai J Med*, 2003. 70(1): p. 3–22. PMID: 12516005
8. Candore G., et al., Biology of longevity: role of the innate immune system. *Rejuvenation Res*, 2006. 9(1): p. 143–8. <https://doi.org/10.1089/rej.2006.9.143> PMID: 16608411
9. Yatsunenko T., et al., Human gut microbiome viewed across age and geography. *Nature*, 2012. 486(7402): p. 222–7. <https://doi.org/10.1038/nature11053> PMID: 22699611
10. Mariat D., et al., The Firmicutes/Bacteroidetes ratio of the human microbiota changes with age. *BMC Microbiol*, 2009. 9: p. 123. <https://doi.org/10.1186/1471-2180-9-123> PMID: 19508720
11. Vemuri R., et al., Gut Microbial Changes, Interactions, and Their Implications on Human Lifecycle: An Ageing Perspective. *Biomed Res Int*, 2018. 2018: p. 4178607. <https://doi.org/10.1155/2018/4178607> PMID: 29682542
12. Odamaki T., et al., Age-related changes in gut microbiota composition from newborn to centenarian: a cross-sectional study. *BMC Microbiol*, 2016. 16: p. 90. <https://doi.org/10.1186/s12866-016-0708-5> PMID: 27220822
13. Biagi E., et al., Through ageing, and beyond: gut microbiota and inflammatory status in seniors and centenarians. *PLoS One*, 2010. 5(5): p. e10667. <https://doi.org/10.1371/journal.pone.0010667> PMID: 20498852
14. Ley R.E., et al., Human gut microbes associated with obesity. *Nature*, 2006. 444(7122): p. 1022–1023. <https://doi.org/10.1038/4441022a> PMID: 17183309
15. Candela M., et al., Maintenance of a healthy trajectory of the intestinal microbiome during aging: a dietary approach. *Mech Ageing Dev*, 2014. 136–137: p. 70–5. <https://doi.org/10.1016/j.mad.2013.12.004> PMID: 24373997
16. Saraswati S. and Sitaraman R., Aging and the human gut microbiota—from correlation to causality. *Frontiers in Microbiology*, 2015. 5: p. 764. <https://doi.org/10.3389/fmicb.2014.00764> PMID: 25628610
17. Bolger A.M., Lohse M., and Usadel B., Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 2014. 30(15): p. 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170> PMID: 24695404
18. Langmead B. and Salzberg S.L., Fast gapped-read alignment with Bowtie 2. *Nature methods*, 2012. 9(4): p. 357. <https://doi.org/10.1038/nmeth.1923> PMID: 22388286
19. Segata N., et al., Metagenomic microbial community profiling using unique clade-specific marker genes. *Nature methods*, 2012. 9(8): p. 811–814. <https://doi.org/10.1038/nmeth.2066> PMID: 22688413
20. Methé B.A., et al., A framework for human microbiome research. *nature*, 2012. 486(7402): p. 215. <https://doi.org/10.1038/nature11209> PMID: 22699610
21. Abubucker S., et al., Metabolic reconstruction for metagenomic data and its application to the human microbiome. *PLoS Comput Biol*, 2012. 8(6): p. e1002358. <https://doi.org/10.1371/journal.pcbi.1002358> PMID: 22719234
22. Buchfink B., Xie C., and Huson D.H., Fast and sensitive protein alignment using DIAMOND. *Nature methods*, 2015. 12(1): p. 59–60. <https://doi.org/10.1038/nmeth.3176> PMID: 25402007
23. Wu C.H., et al., The Universal Protein Resource (UniProt): an expanding universe of protein information. *Nucleic acids research*, 2006. 34(suppl_1): p. D187–D191. <https://doi.org/10.1093/nar/gkj161> PMID: 16381842
24. Ye Y. and Doak T.G., A parsimony approach to biological pathway reconstruction/inference for genomes and metagenomes. *PLoS Comput Biol*, 2009. 5(8): p. e1000465. <https://doi.org/10.1371/journal.pcbi.1000465> PMID: 19680427
25. Venables WN, Ripley B., *Modern applied statistics with S. Statistics and computing*. New York: Springer, 2002.
26. Kassambara A. and Mundt F., *Factoextra: extract and visualize the results of multivariate data analyses*. R package version, 2017. 1(5): p. 337–354.
27. Anderson M.J., A new method for non-parametric multivariate analysis of variance. *Austral ecology*, 2001. 26(1): p. 32–46.
28. Morgan X.C., et al., Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome biology*, 2012. 13(9): p. R79. <https://doi.org/10.1186/gb-2012-13-9-r79> PMID: 23013615
29. Leisch F. and Dimitriadou E., *Machine Learning Benchmark Problems*. R Package, mlbench, 2010.
30. Kuhn M., *caret: Classification and Regression Training*, Version: 6.0–82. Retrieved April 16, 2019.

31. Breiman L., Random forests. *Machine learning*, 2001. 45(1): p. 5–32.
32. Sing T., et al., ROCR: visualizing classifier performance in R. *Bioinformatics*, 2005. 21(20): p. 3940–3941. <https://doi.org/10.1093/bioinformatics/bti623> PMID: 16096348
33. Ahdesmäki M. and Strimmer K., Feature selection in omics prediction problems using cat scores and false nondiscovery rate control. *The Annals of Applied Statistics*, 2010. 4(1): p. 503–519.
34. Supek F., et al., REVIGO summarizes and visualizes long lists of gene ontology terms. *PloS one*, 2011. 6(7): p. e21800. <https://doi.org/10.1371/journal.pone.0021800> PMID: 21789182
35. Arumugam M., et al., Enterotypes of the human gut microbiome. *nature*, 2011. 473(7346): p. 174–180. <https://doi.org/10.1038/nature09944> PMID: 21508958
36. Landis J.R. and Koch G.G., The measurement of observer agreement for categorical data. *biometrics*, 1977: p. 159–174. PMID: 843571
37. Cresci G.A. and Bawden E., Gut Microbiome: What We Do and Don't Know. *Nutr Clin Pract*, 2015. 30(6): p. 734–46.
38. Shin N.R., Whon T.W., and Bae J.W., Proteobacteria: microbial signature of dysbiosis in gut microbiota. *Trends Biotechnol*, 2015. 33(9): p. 496–503. <https://doi.org/10.1016/j.tibtech.2015.06.011> PMID: 26210164
39. Rizzatti G., et al., Proteobacteria: A Common Factor in Human Diseases. *Biomed Res Int*, 2017. 2017: p. 9351507. <https://doi.org/10.1155/2017/9351507> PMID: 29230419
40. Zysset-Burri D.C., et al., Retinal artery occlusion is associated with compositional and functional shifts in the gut microbiome and altered trimethylamine-N-oxide levels. *Scientific reports*, 2019. 9(1): p. 1–11. <https://doi.org/10.1038/s41598-018-37186-2> PMID: 30626917
41. Azad M., et al., Probiotic species in the modulation of gut microbiota: an overview. *BioMed research international*, 2018. 2018. <https://doi.org/10.1155/2018/9478630> PMID: 29854813
42. Jie Z., et al., The gut microbiome in atherosclerotic cardiovascular disease. *Nat Commun*, 2017. 8(1): p. 845. <https://doi.org/10.1038/s41467-017-00900-1> PMID: 29018189
43. Li J., et al., Gut microbiota dysbiosis contributes to the development of hypertension. *Microbiome*, 2017. 5(1): p. 14. <https://doi.org/10.1186/s40168-016-0222-x> PMID: 28143587
44. Huynh-Thu V.A., et al., Statistical interpretation of machine learning-based feature importance scores for biomarker discovery. *Bioinformatics*, 2012. 28(13): p. 1766–1774. <https://doi.org/10.1093/bioinformatics/bts238> PMID: 22539669
45. Foster K.R., Koprowski R., and Skufca J.D., Machine learning, medical diagnosis, and biomedical engineering research-commentary. *Biomedical engineering online*, 2014. 13(1): p. 1–9.
46. Mohan S., Thirumalai C., and Srivastava G., Effective heart disease prediction using hybrid machine learning techniques. *IEEE Access*, 2019. 7: p. 81542–81554.
47. Magne F., et al., The Firmicutes/Bacteroidetes ratio: a relevant marker of gut dysbiosis in obese patients? *Nutrients*, 2020. 12(5): p. 1474. <https://doi.org/10.3390/nu12051474> PMID: 32438689
48. Zysset-Burri D.C., et al., Associations of the intestinal microbiome with the complement system in neovascular age-related macular degeneration. *NPJ genomic medicine*, 2020. 5(1): p. 1–11.
49. Walle P., et al., Alterations in fatty acid metabolism in response to obesity surgery combined with dietary counseling. *Nutrition & diabetes*, 2017. 7(9): p. e285–e285. <https://doi.org/10.1038/nutd.2017.33> PMID: 28869586
50. Molano A., et al., Age-dependent changes in the sphingolipid composition of mouse CD4+ T cell membranes and immune synapses implicate glucosylceramides in age-related T cell dysfunction. *PLoS One*, 2012. 7(10): p. e47650. <https://doi.org/10.1371/journal.pone.0047650> PMID: 23110086
51. Rienzi S.C.D., et al., The microbiome affects liver sphingolipids and plasma fatty acids in a murine model of the Western diet based on soybean oil: Hepatic sphingolipids and plasma FAs are altered by gut microbes. *J Nutr Biochem*, 2021: p. 108808. <https://doi.org/10.1016/j.jnutbio.2021.108808> PMID: 34186211
52. Atamna H., Heme, iron, and the mitochondrial decay of ageing. *Ageing Res Rev*, 2004. 3(3): p. 303–18. <https://doi.org/10.1016/j.arr.2004.02.002> PMID: 15231238
53. Wang R., et al., Four-week administration of nicotinemoderately impacts blood metabolic profile and gut microbiota in a diet-dependent manner. *Biomed Pharmacother*, 2019. 115: p. 108945. <https://doi.org/10.1016/j.biopha.2019.108945> PMID: 31100541