A Study on the Analysis of Personal Gut Microbiomes

Journal of
Computer Science & Systems Biology

Research Article Open Access

A Study on the Analysis of Personal Gut Microbiomes

Olga Vasieva[1#], Anatoly Sorokin[2], Marsel Murzabaev[3], Peter Babiak[3] and Igor Goryanin[3,4,5*#]

[1]*Ingenet Ltd., UK*
[2]*BMI Sciences Ltd., UK*
[3]*Okinwawa Institute Science and Technology, Japan*
[4]*School of Informatics, University of Edinburgh, UK*
[5]*Tianjin Institute of Industrial Biotechnology, China*
[#]*Equal contribution*

Abstract

In this case study, we validate a personalized approach to gut microbiome (GM) analysis and interpretation based on published association studies. We apply our ASAR data annotation and clustering package to a series of 10 sequenced GM's from individuals of different ethnical and geographical backgrounds, age and health groups. The differentially presented and detectable taxonomic and functional signatures in each GM metagenome are used to predict the hosts' characteristics via correlations established in published studies, and the predictions are being validated by available individual-associated metadata. We also test sensitivity of the routine annotation and data clustering pipeline to an individual and family-linked signatures in GM structure and functionalities, when applied to a limited number of varying samples. The number of samples was sufficient to demonstrate 2 main types of GM composition, based on Bacteroides or *Prevotella* as main abundant genera, limitation of a variety of taxa as a result of antibiotics application, clustering of family members' GM metagenomes both in taxonomic and in functional space, individual signatures related to chronic diseases and pharmacological interventions, and elements of ethnicity-related characteristics in the metagenomes. The method and logical algorithms of the analysis applied here may be utilized in rather computational pipeline for a personalized microbiome analyses, and their potential useful outputs and limitations are being discussed.

Keywords: Gut microbiome; Metagenome; Personalized; Diagnostic; Therapies; Electrogenic treatment; ASAR visualization tool; MGRAST server

Introduction

Interactions between multicellular organisms and their environments are largely transmitted via associated prokaryotes providing beneficial as well as damaging chemical factors. A digestive system is an environmental frontline involving digestive secretions, intestinal cell metabolism and signalling and the gut microbiome (GM), where the latter adapts to an individual's lifestyle, gut environment changing pathologies and pharmacological interventions [1,2] and also significantly modulate them [3]. Nutrients and prebiotics provide substrates for a dynamic GM which is estimated to consist of over 1000 different microbial species belonging to five predominant phyla: *Firmicutes, Bacteroidetes, Actinobacteria, Verrucomicrobia* and *Proteobacteria* [4-7]. The about 400 identified species are strictly anaerobic and hence will generally be found in mucosal regions such as the oral cavity and the GI tract [2,6].

GM are shown to affect our behaviour, cognitive functions and emotions, and, literally, make us. Two thirds of each microbiome were suggested to be individual-specific 'fingerprint' that may tell us a lot about an individual [2,8]. Taking in mind that each GM is comprised by about 3 million genes and there are particular correlations between GM and diet [1,9-11], geography and ethnicity of a host [12-15], age and longevity [10,16,17], and certain diseases [8,18-22]. However, it is still a challenge to define and interpret individual's GM characteristics [23-25], where a selection of a control group seems to be the main challenge. We still need to be able to define a 'healthy individual's spectrum of GM composition variations, to contrast the latter to not only any well manifested disease but to rather a pre-diseased state of a host. That will lead to a use of GM analysis as a part of the preventive medicine approach, and diagnostics. We also need to integrate more data to be able to interpret taxonomic and functional 'signatures' of a particular GM as many links from bacteria in a gut community to GM metabolic flows, metabolite absorption and host's metabolism still remain a mystery.

By this small case study, we validate a personalized approach to GM analysis and interpretation. As an extensive research has been targeting potential correlations between diets, genetic background and gut microbiome's composition, we rather try to apply the accumulated knowledge. We use our ASAR annotation and clustering analysis tool [26] on a series of ten sequenced GM's from individuals of different ethnical and geographical backgrounds, age and health groups. By manual analysis of the defined individual signatures we've reconstructed connections from their exclusive taxonomic and differential metabolic capabilities to potential hosts' characteristics to highlight the power and limitations of the approach. Most of metadata was not included in consideration until the post-analytic stage and have been used for validation of the predictions made from individual's GM-specific taxonomic and metagenomic functional characteristics. A limited number and a diversity of bacterial compositions of the GM samples used in this study, presented a case that can be typical for a clinical setting and also highlighted a need for rather flexible stratification approach in selection of specific contrasting datasets for personalized GM analysis to avoid complex family, ethnic and geographical biases. The results of our case study may help in design of new algorithms for selection of such contrasts and pipelines for a computerized GM-based preventive medical diagnostics.

***Corresponding author:** Igor Goryanin, Okinwawa Institute Science and Technology, Japan, Tel: 00966539417360; E-mail: goryanin@oist.jp

Received July 18, 2019; **Accepted** September 05, 2019; **Published** September 13, 2019

Citation: Vasieva O, Sorokin A, Murzabaev M, Babiak P, Goryanin I (2019) A Study on the Analysis of Personal Gut Microbiomes. J Comput Sci Syst Biol 12: 71-79.

Methods

Sample preparation

The samples were taken using *OMNI* gene GUT/OM-200 kits (DNA-Genotek, Canada). We also tested if off-shelf treatment related to microbial bioelectricity (electrogenic treatment) may help to access the low-abundant GM bacterial genera. Stool samples from 3 individuals were also taken before and after an intake of 10 capsules each containing 170 mg electrogenic treatment (purchased from Kannabe Hakutan Kobo, Hyogo, Japan) indicated in Table as /1 for a day on electrogenic treatment, /2 for a day after a electrogenic treatment and /3 for the next day sample.

Whole genome sequencing

DNA was extracted, general DNA Library constructed and sequenced on BGISEQ PE100. 4GB of data were received for every metagenome.

Data processing

The sequencing data were uploaded to the MG-RAST server as FASTAQ files for processing, primary analysis, and storage (Figure 1). *H. sapience* (human) genome sequences were marked for exclusion during data submission. Primary submission data and results of the MG-RAST pipeline are available publicly. The MG-RAST representative hit organism abundances calculation was performed against the SEED database at the level of genera, based on a maximum *e*-value of 1×10^{-5}, minimum identity cut-off of 60%, and minimum sequence alignment of 15. Abundance data were downloaded as TSV files for further analysis. The representative hit data were downloaded from MG-RAST server via MGRASTer package [27] in R 3.1 environment. Abundance analysis was performed in metagenome Seq package [28] and ordination analysis was performed with phyloseq R packages [29]. Krona taxonomic community profiles were built by MG-RAST and stored as an image.

Data visualization

Functional, taxonomic, and KEGG orthology data were obtained from reads via MG-RAST pipeline. The functional and taxonomic annotations were merged based upon identical md5's corresponding to unique read sequences. Then read counts were aggregated for reads annotated with the same function and taxon. Functional and taxonomic read annotations to lowest level are matched to the lowest level annotations in their corresponding hierarchy trees to generate the whole phylogeny of each read.

Our post-annotation analysis and visualization tool, ASAR (Figure 1) [26] uses data integration algorithm to merge taxonomic and functional data annotated at read level. The resulting 3D datasets with axes of Functions, Taxonomy and Metagenome samples were visualized via three heatmaps of each axis versus two others (F&T, F&M, T&M). Additionally, KEGG pathway enrichment sorting/heatmap and its map visualization were implemented. Advantages of the tool are: 1) Integrated functional and taxonomic analysis; 2) Comparative analysis of pathway enrichments; 3) KEGG pathway map visualization. The heatmaps show log abundance of reads annotated with selected functions in particular taxa within particular communities. On the KEGG map each functional box is split into sections corresponding to analyzed bacterial communities. The relative abundance of each function in each community is color coded from green (the lowest) to dark red (the highest proportion in the community).

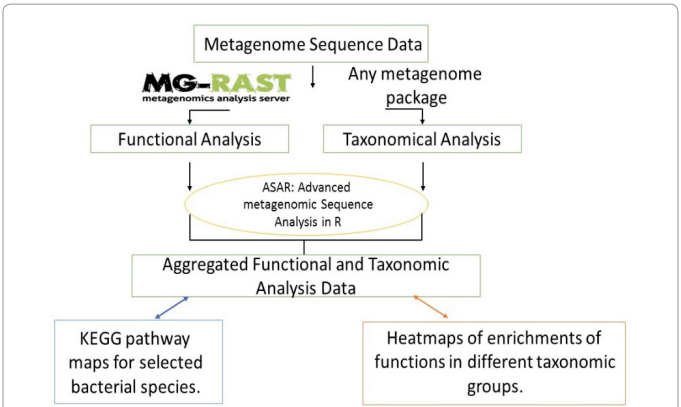

Figure 1: A schematic presentation of an ASAR-based metagenome analytical pipeline.

Results and Analysis

Taxonomic analysis

The methods of stool sampling, sequencing and analysis used in this study lead to a consistent detection of at least 9 of 10 top abundant genera from each individual's GM after 3 repeated feces collections. To check if there were biases in presentation of any narrowed gut compartment in each individual's feces, we've used an off-shelf treatment related to microbial bioelectricity as a perturbation of a feces reaching microbiome composition. As result of the intake, Individual-specific changes in ranking of the most abundant bacterial taxa were noticed, with *Prevotella* and Bacteroides dominance increasing after the intake (Table 1, Individuals 5, 6, 7). Some genera, as *Oscillibacter* show consistent decline after the treatment, though it doesn't change the ranking of the bacterial group in the top 10 genera.

Relative percentage abundancies of individual's top 10 GM bacterial genera in feces samples (Table 1). To estimate a consistency of a sampled GM composition, feces from 3 individuals were also collected straight after an intake of off-shelf supplementary medicine and a day after indicated in Table as /1 for a day before treatment, /2 for a day after a electrogenic treatment and /3 for the next day sample.

Oncoming publication and here we would like to rather stress the fact of a relevant robustness of the individuals' 5,6 and 7 GM structure. Interestingly, the top genera repeatedly provide 50-60% of the total GM DNA content with only one exception (34%-sample 6/1, which was brought to 58% after the electrogenic treatment). Several noticed correlations are well in accordance with the published observations. Individual GMs occurred to be Bacteroidetes family genera of either *Prevotella* or Bacteroides dominating, with only one exception (Individual 3), where high abundancies of both *Prevotella* (27.9%) and Bacteroides (16.5%) were balanced by a decreased abundancy of *Faecalibacterium* (3.2%). Individual 7's signature in low content of both *Prevotella* (3.5%) and Bacteroides (2.3%) was compensated by sudden and progressive increase in *Prevotella* abundance after the electrogenic treatment. Both Bacteroides and *Prevotella* show a clear tendency to expand in abundancy after the electrogenic treatment that may be indicative of their particular attachment or unavailability for electrogenic treatment particles due to a specific geography of location of these genera in the gut [30]. Interestingly, *Roseburia* also shows similar tendency and suggestively also shares a niche with the most abundant GM residents. Firmicute *Roseburia* is a beneficial butyrate-producing bacteria and a primary degrader of dietary beta-mannans

***Genus*\Individual**	**1**	**2**	**3**	**4**	**5/1**	**5/2**	**5/3**	**6/1**	**6/3**	**7/1**	**7/2**	**7/3**
Faecalibacterium	13.7	6.4	20.6	3.2	14.4	19.6	11.4	15.1	18.9	13.3	14.5	10.1
Prevotella	-	-	13.0	27.9	-	-	-	9.1	11.6	3.5	17.6	26.4
Alistipes	2.6	6.2	7.4	2.2	12.1	13.0	9.3	15.9	11.6	4.0	4.8	3.1
Ruminococcus	-	-	5.8	-	-	-	-	1.7	1.9	2.6	-	1.1
Bacteroides	32.4	41.5	3.1	16.5	16.4	16.0	29.6	3.0	4.8	2.3	4.0	8.2
Oscillibacter	-	0.4	1.7	1.2	2.2	1.7	1.0	1.5	1.0	3.8	3.3	1.5
Sutterella	1.4	-	1.3	1.7	1.6	3.4	1.5	-	-	1.5	2.9	1.0
Blastocystis	-	-	0.9	-	-	-	-	-	-	-	-	-
Clostridium	1.8	1.7	0.9	1.5	4.2	3.4	2.7	5.6	6.1	1.8	2.0	1.1
Parabacteroides	1.3	0.6	0.8	2.1	0.8	1.1	1.4	-	-	-	-	0.8
Bifidobacterium	-	-	-	-	1.4	-	-	1.8	1.2	-	-	-
Bilophila	-	-	-	-	1.0	0.9	1.1	-	-	1.1	1.4	-
Paraprevotella	-	-	-	1.0	0.9	0.9	1.3	-	-	-	-	-
Roseburia	1.2	0.6	-	-	-	0.9	1.0	-	-	-	-	1.3
Desulfovibrio	-	-	-	1.3	-	-	-	-	-	2.0	2.0	-
Collinsella	-	-	-	-	-	-	-	-	-	-	0.9	-
Akkermansia	-	-	-	-	-	-	-	1.6	3.4	-	-	-
Flavonifractor	-	-	-	-	-	-	-	1.3	0.8	-	-	-
Streptococcus	1.4	-	-	-	-	-	-	-	-	-	-	-
Escherichia	0.6	-	-	-	-	-	-	-	-	-	-	-
Odoribacter	0.4	-	-	-	-	-	-	-	-	-	-	-
Phascolarctobacterium	-	0.5	-	-	-	-	-	-	-	-	-	-
Eubacterium	-	0.2	-	-	-	-	-	-	-	-	-	-
Lachnoclostridium	-	0.2	-	-	-	-	-	-	-	-	-	-

Table 1: Relative percentage abundancies of individual's top 10 GM bacterial genera in feces samples. To estimate a consistency of a sampled GM composition, feces from 3 individuals were also collected straight after an intake of off-shelf supplementary medicine and a day after indicated in Table as /1 for a day before treatment, /2 for a day after a electrogenic treatment and /3 for the next day sample.

[31]. Butyrate produced by commensal bacteria serves as the main energy source for colonocytes and it exhibits anti-carcinogenic and anti-inflammatory properties in the distal gut [32-34], that makes our finding applicable to beneficial modulation of the composition of the gut microbiota.

There are differential taxonomic compositions signatures associated with each host that may be attributed to particular health characteristics and, therefore, used as predictive markers. First of all, compositions of GM of individuals 5 and 6 can be characterized as healthy by presence of such probiotics as *Bifidum*, *Roseburia* (6) and high levels of *Alistips* (>10% av.) with a high levels of *Faecalibacterium* (>10% AV) associated with *Prevotella* (6) or rather Bacteroides (5) basic trends. GM of the 3d individual is not enriched by particular probiotics but it an illustration of the balanced GM composition, associated with the healthy trends according to a number of studies [1,8,18-22,35]. A limited number of samples used in our analysis was sufficient to illustrate an existence of gut GM 'community types' [23], identifiable by variations in the level of one of three genera: Bacteroides (Enterotype 1), *Prevotella* (Enterotype 2) and *Ruminococcus* (Enterotype 3) [24] but found somehow controversial, as thoroughly reviewed [25].

Based on *Prevotella* prevalence we can attribute 3,4,6,7 GMs to rather vegetarian diet health styles or particular ethnicities [13]. The defined signatures would be in correspondence with the ethnicity of the individuals [5-7,36]. Interestingly, high levels of *Alistipes* were shown to be associated with a particular mucosa composition determined by secretor (FUT2) genotype [37]. Taking in mind a high frequency of this genotype in the Asian population the high abundance of *Alistipes* and comparably elevated Bacteroides abundancy level in GM 5 may be considered as a reflection of the particular genotype manifestation. GM composition of individual 6 is particularly based on a strong presence of *Alistipes* (> 15%) and is also characteristic by a presence of a probiotic mucin-degrader as *Akkermansia* [38].

There are also alarming microbiome composition signatures that may be taken in consideration. The 4th and 7th GMs show a presence of electrogen *Desulfovibrio*, which has been associated with particular host's neural modulation via production of toxic sulphides [21,22]. However, both individuals are generally healthy and do not demonstrate any associations with the mentioned diseases. It is a good demonstration of a difficulty of extrapolation of statistically significant correlations shown on large cohorts on each particular individual's case.

The 2nd GM's composition is very limited, with a demonstrated striking Bacteroides domination (41.5%). The corresponding individual is under periodic prophylactic antibiotics treatment over the last 5 years, which is in a good agreement with the poor taxonomic diversity of the GM. As in case with a electrogenic treatment, domination of mucus-adherent genera [39] can be caused by a selective elimination of the bacteria that are more accessible to the intervening medication. *Faecalibacteria* and *Alistipes*, distal gut bacterial genera show moderate abundancies (> 6%). The 1st GM's host, who is a biological father to the host 2 is also characterized by a strong domination of Bacteroides genera (32%). Such a correlation makes us to suggest that there may be a role of genetic factors benefiting the environment for Bacteroides propagation [37]. Functional competition between bacteria for existence in the mucus layer is suggested to be a major determinant of the sustained microbiota composition within the host and may be strongly affected by the mucus structure [2,37,40].

GM of the individual 1 has also a high level of abundance of *Streptococcus* (1.2%), that could be indeed eliminated by a prophylactic antibiotic in the individual 1's family member, individual 2. Interestingly, *Streptococcus* was the main target of that prophylactics, though a potential association of *Streptococcus* presence with a gut has not been suggested or discussed. We are keeping a time series of samples from individual 2 for further analysis and more support to

our hypothesis of potentially gut origin of Streptococcus constant immunization source.

Relative high abundancy of *Streptococcus* in individual 1's GM accompanied by a comparably high level of *Escherichia* may be linked to specific health aspects of faulty liver cholesterol metabolism: fatty liver and heart pathologies [2]. It has been explained by a sensitivity of *Streptococcus* and *Escherichia* to bile secreted into the proximal gut [41-44]. Bile acids, secreted through the bile duct at the proximal end of the small intestine, are bactericidal to certain species due to their surfactant properties and are known to broadly shape the composition of the microbiota, especially in the small intestine. For example, feeding mice excess bile acids generally stimulates the growth of Firmicutes and inhibits Bacteroidetes [45,46]. There is indeed a cholesterol metabolism and heart dysfunctions diagnosed for the individual 1, which is a potentially very striking fact of the explicit manifestation of a correlation defined via a population study.

Increased abundancy of Lactobacillus and Enterobacteriaceae genera and a depleted number of *Roseburia* group may indicate a decreased butyrate and propionate production in GM of the individual 1 [3,43,47]. Actinobacteria e.g. *Eubacterium*, detected in the GM of individual 2, however, contribute significantly to butyrate production in the colon both directly and via metabolic cross-feeding [48], which makes GM of individual 2 more adherent to a healthy trait. Increased Enterobacteriaceae abundancy may be also signal of iron deficiencies [43,44], which, however, are not diagnosed at the time of this analysis for the individual 1.

Samples were distinguishable at different levels of taxonomic analysis (Figure 2). Sample 1 signature features, at different taxonomic levels, are shown in frames. Some genera from the strong signature at a level of class classification (Figure 2A) were below the top 50 rank threshold, which led to a dissipation of the signature cluster at the lower taxa levels (Figures 2B and 2C). Both class and genera level analysis were shown to be the most informative in defining an individual's GM-specific features.

Functional analysis of GM metagenomes

Metagenomics allows analysis of not only taxonomic composition of a GM but also of genomic capabilities attributed to metabolic and virulent features of the bacteria comprising the community. More similarities were observed between the functionality profiles of microbial genes present in repeat samples from same individual's GM than between their taxonomic profiles, suggesting that the core microbiota may be better defined at a functional rather than taxonomic classification level [49]. However, analysis of 50 most presented in each GM metagenome metabolic functions showed clear differences between the samples from different individuals (Figure 3). Despite the variety of the dominant taxa (Table 1) there is a clear clustering of the metagenomes from the family members' GMs at the level of their functional capabilities, which is comparable to a clear co-clustering of metagenomes from repeat samples (7/1, 7/3). For more details please refer Supplementary Figure 3.

The family members' GMs were found to share a specific increase in genomic presence of glycosidases function, fucosidase and galactosidase in particular. Fucose is a major component of human mucin glycoproteins and glycolipids and is also present in foods and may be affected by milk-reach diet [50]. However, a presence of milk-degrading probiotic bacteria from the *Bifidum* genera is, on contrary, diminished in GMs from individuals 1 and 2 (Table 1). This functional signature thus can be attributed to a high abundancy of mucus-degrading representatives of Bacteroides genus (Table 1, Figures 2 and 3), typical for individuals 1 and 2, and may have a certain ethnicity background [B]. Destruction of the mucosal glycoproteins is one of the factors leading to increased permeability of mucus for bacteria that are not normal mucosal residents, such as *E. coli*, with consequent inflammation of endothelium of the gut and formation of ulcers [44,51-53]. As there is a potential hazard in the propagation of these functions, especially for an individual 1, specific recommendations for a diet and a health check could be made. Interestingly GM-1 metagenome shows a clear aerobic signature of particularly abundant genomic functional features (Figures 4 and 5). Aerobicity of the gut environment is an indication of vascularization of the intestinal wall [40] which often corresponds to ongoing inflammation. In compare to all the other GM samples in terms of taxonomy (increased abundancy of *E. coli*) and functional capacity (increased abundancy of mucus degrading functions), the indicated acrogenic/oxidative environment would be counted as a negative prognosis component [2,40]. As it was mentioned above, there is also the individual 1 GM's attributed

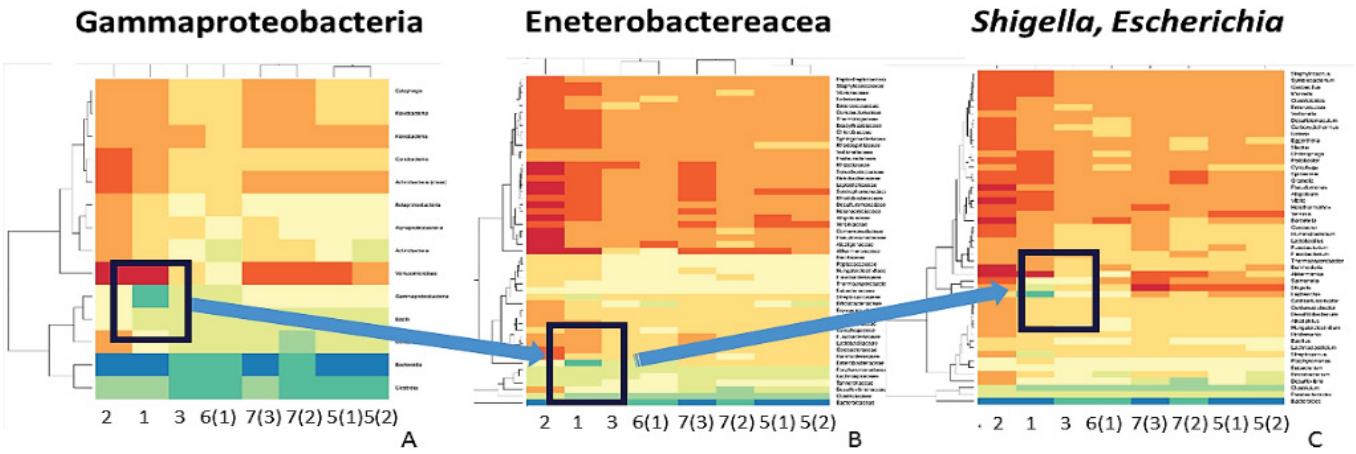

Figure 2: Individual GM metagenomes (indicated by horizontal axis numbers in accordance with Table 1) were distinguishable at different levels of taxonomic analysis. Sample 1 signatures are shown in frames and indicated above the heatmaps. Top 50 taxa were shown in heatmaps B and C. A heatmap block colour corresponds to mapped sequences abundancies in the rainbow order: from dark red-the lowest to dark blue-the highest.

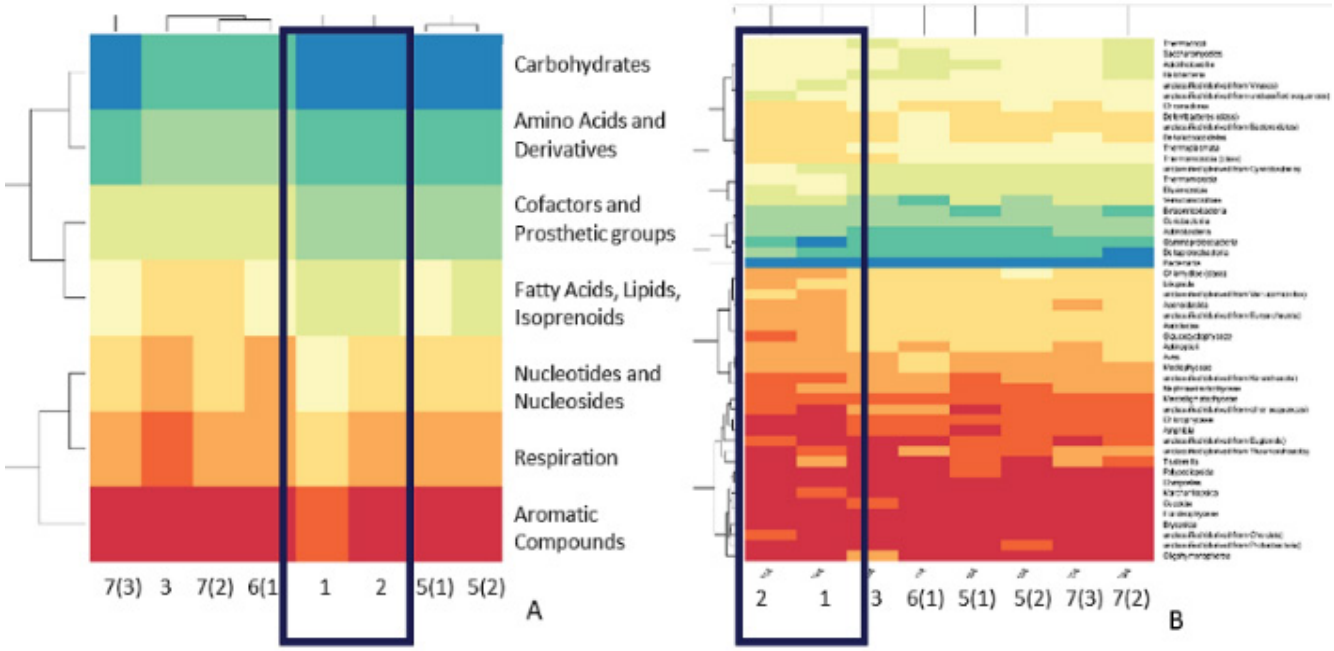

Figure 3: Heatmap illustrating clustering of the functional metagenomics features attributed to the GMs from 1-7 individuals (as described in Table 1). A-general metabolic categories, B-metabolic subsystems (SEED). Columns attributed to 2 members of one family are framed A heatmap block color corresponds to mapped sequences abundancies in the rainbow order: from dark red-the lowest to dark blue-the highest. For details see Supplementary Figures 1 and 2.

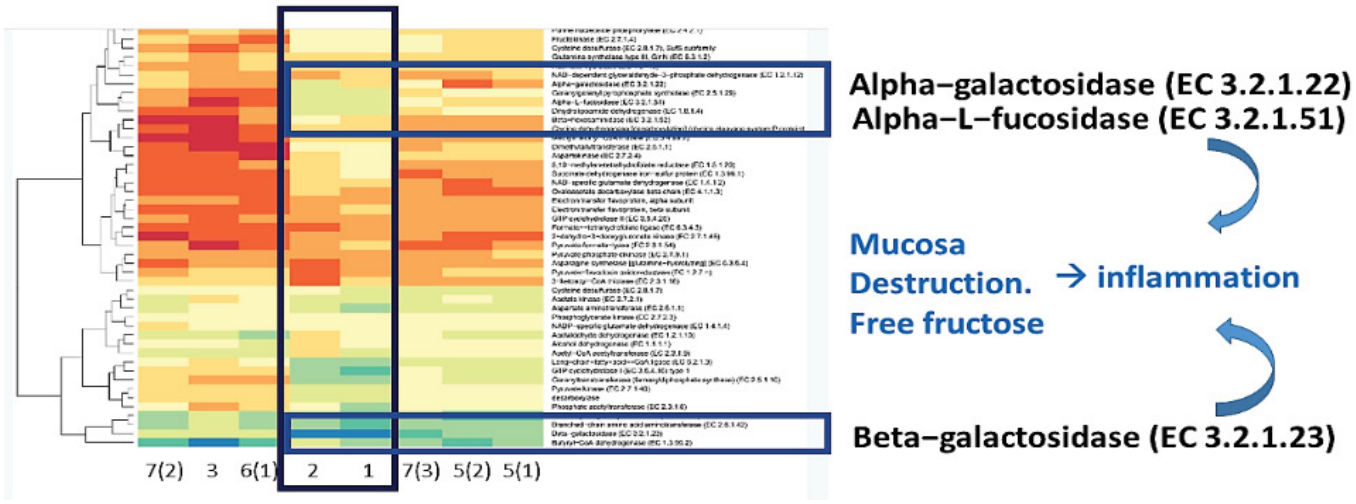

Figure 4: Heatmap illustrating clustering of the functional metagenomics features attributed to the GMs from 1-7 individuals (as described in Table 1). Sugar hydrolase functions corresponding to blocks of high abundancy sequences (framed) in GMs of individuals 1 and 2 (members of one family), are listed. A heatmap block color corresponds to mapped sequences abundancies in the rainbow order: from dark red-the lowest to dark blue-the highest. For more details see Supplementary Figures 1 and 2.

signature indicative of a faulty cholesterol metabolism/liver function that may be also causally relevant to gut inflammation.

An abundance of the most of top 50 functional signatures of individual 1's metagenome has corresponded to the abundancy of the taxa possessing the particular functionality (such as Dihydroneopterin-triphosphate epimerase typical for E. coli and Shigella of all the Enterobacteriaceae) [54] (Figure 6). Increased abundancy of genomic sequences for Hydroxymethylglutaryl-CoA synthase (HMGS) (Figure 6), enzyme in a mevalonate pathway of isoprenoid biosynthesis, needed some other explanation as it is not specifically associated with the dominating taxa. Human HMGS is the target of STATINS [55,56] acting to decrease rates of cholesterol biosynthesis, and a drug from this group, as a matter of fact, is prescribed to the individual 1. We have hypothesized that intake of STATINS could lead to a selection of the gut bacteria strains possessing increased number of the HMGS encoding genes to compensate for the bacterial HMGS inhibition. The latter may be responsible for the anti-bacterial effect of STATINS [57]

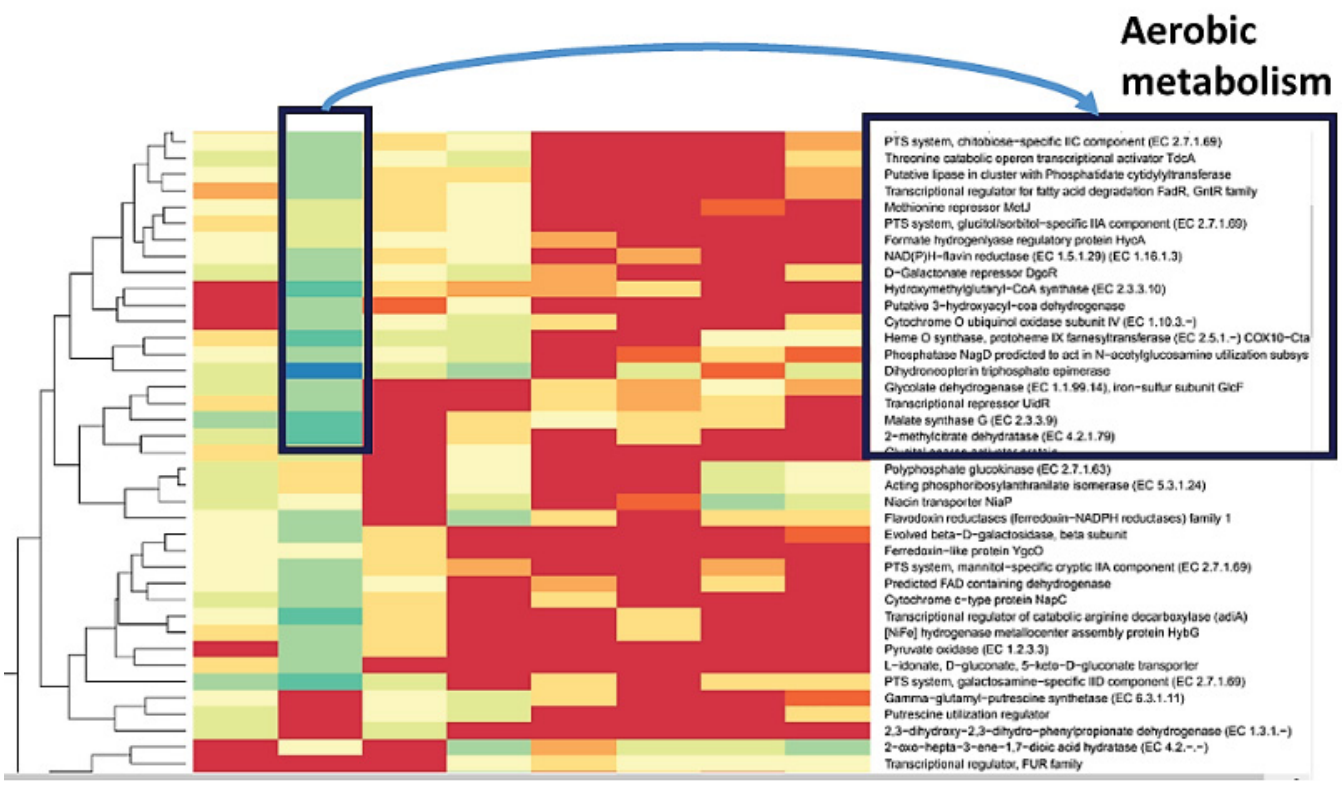

Figure 5: Heatmap illustrating clustering of the functional metagenomics features attributed to the GMs from 1-7 individuals (as described in Table 1). A heatmap block color corresponds to mapped sequences abundancies in the rainbow order: from dark red-the lowest to dark blue-the highest. The individual 1's GM metagenomics functional signature is highlighted.

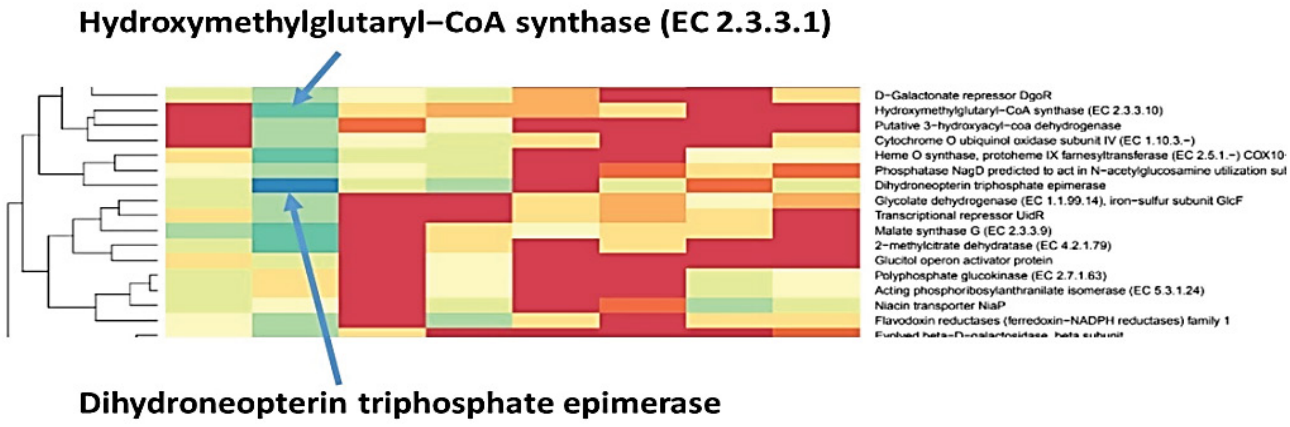

Figure 6: Part of the Heatmap (Figure 5) illustrating clustering of the functional metagenomics features attributed to the 1-7GMs (as described in Table 1). The functional signatures of individual 1's GM are highlighted. A heatmap block colour corresponds to mapped sequences abundancies in the rainbow order: from dark red-the lowest to dark blue-the highest. For more details see Supplementary Figures 1 and 2.

but the mechanism of this action still requires detailed investigation. Other plausible explanation may be that over presentation of the taxa able to synthesis mevalonate could be a way to compensate for the diminished amount of HMGS-downstream products normally synthesized and provided by the host.

A unique profile of the GM metagenome of individual 1 can be also attributed to other pharmacological-interventions, such as by non-steroidal anti-inflammatory drugs [58] regularly taken by the individual 1. As the potential drug-linked GM profiles and predictors overlap, correlation between GM community structure/function and particular drugs is still not predictive and requires further investigation on a large number of cases.

Associated virulent factors show an increased abundancy in GM 1 metagenome (Figure 7), that in general may reflect the discussed factors that affect interactions between a host and its gut microbiome. A very indicative and a unique feature of an individual GM is also its

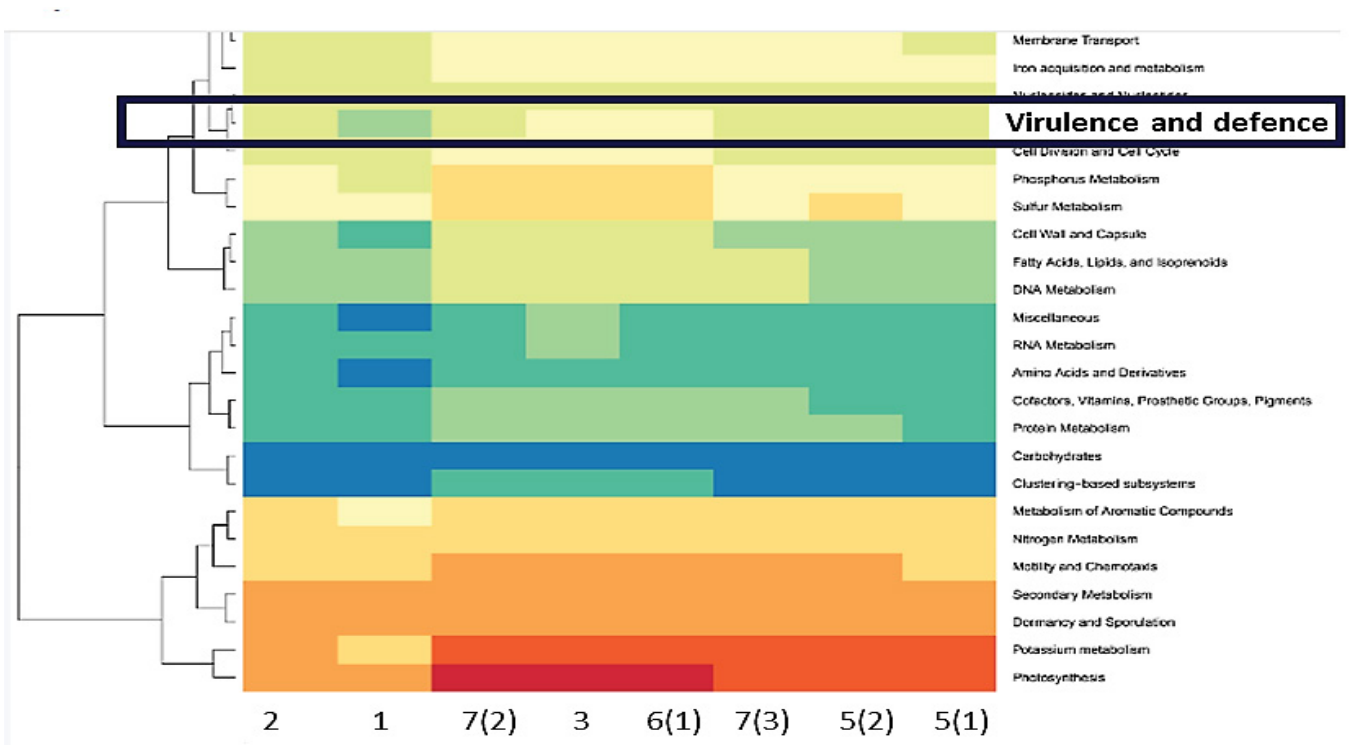

Figure 7: Abundancies of sequences associated with 'Virulent factors' category in the metagenomes. A heatmap block colour corresponds to mapped sequences abundancies in the rainbow order: from dark red-the lowest to dark blue-the highest.

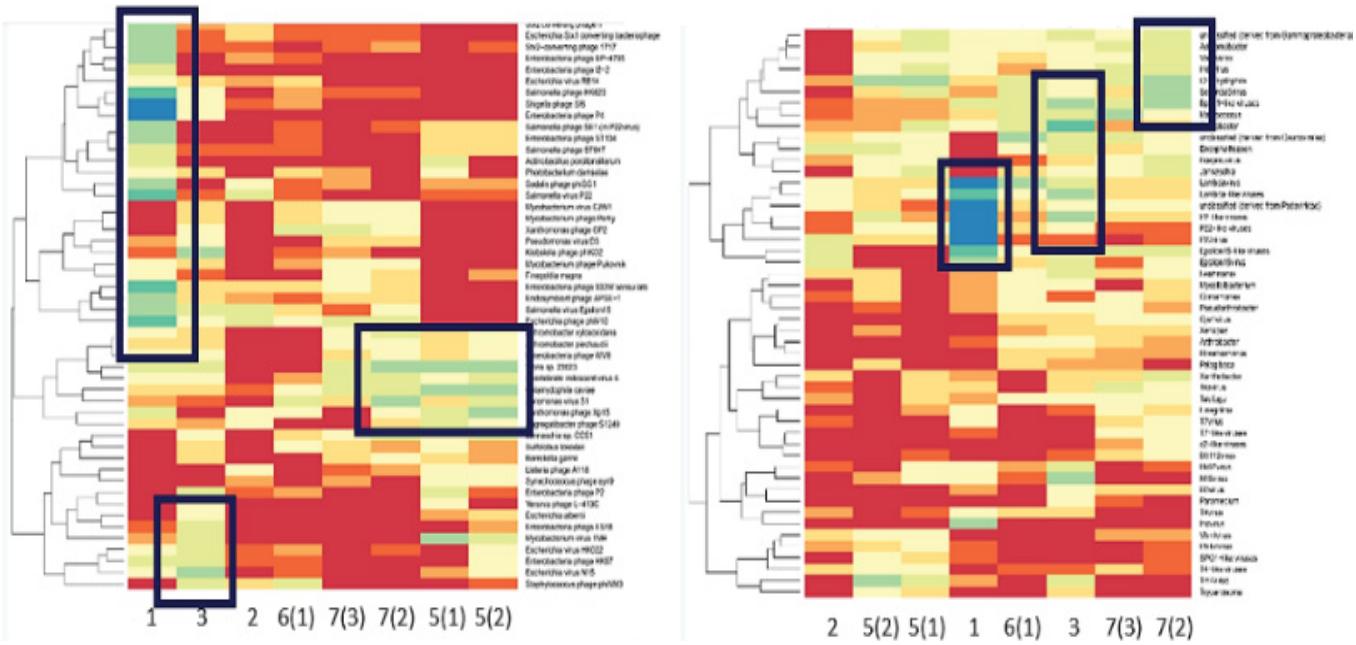

Figure 8: Abundancies of sequences mapped to bacteriophages from the analysed GM metagenomes. The examples of signature clusters are shown in frames. A heatmap block colour corresponds to mapped sequences abundancies in the rainbow order: from dark red-the lowest to dark blue-the highest.

bacteriophage profile which likely follows the geographical and ethnical connections (Figure 8). The uniqueness of these attributes can be used to investigate connections between individuals in social environments.

Discussion

Given the complex relationship existing between the gut microbiota

and the host, it is not surprising to observe a divergence from the normal microbiota composition in individual GMs. Given the contribution of host genetics in many diseases associated with a dysbiotic microbiota, dual preventive strategies (targeting not only a host, but also its GM) may be required to restore the physiological balance. However, we still need to understand where dysbiosis starts. Can we suggest its presence solely from the analysis of the individual's microbiota composition and known significant correlations at a population level? These correlations are subjected to particular quantitative ratios between all bacteria in the individual GM community and specific characteristics (genetics, ethnicity, age) of a healthy host [8-17], which we cannot yet see in their full integrative crosstalk.

From this study we can, however, advice on a design of new algorithms for selection of contrasts and pipelines for a potentially computerized GM-based preventive medical diagnostics. Taxonomic analysis allows comparison of individual GM composition with GM compositions shown to be significantly linked to particular diets and, in some cases, chronic diseases [1,9-11,14-17]. More statistically grounded analysis is required to allocate an individual microbiome in the space of GM compositions associated to potential health hazards and to avoid biases linked to ethnicity, geography and family diet trends that may have different manifestations in hosts with different backgrounds.

Cherry-picked individual's GM signatures may not necessary correspond to significant correlations described in literature (as in the case of *Desulfovibrio* abundancy/autism link [21] not observed in our study), however, they may indicate a presence of certain hidden health issues. Reversed correlative analysis (from GM's taxonomic and functional signatures to associated clinical data) should probably take place to accompany the usual approach centred at patients' health/clinical characteristics and patient contrasting groupings [1-22]. There is a number of gut microbiome markers [2,38,40,43,44], predicted via this approach, which need to be validated as indeed predictive in personalized preventive medicine.

Family-shared GM composition and the metagenome's functional enrichment signatures can be suggested as baselines for a discovery of an individual health-related GM modulations. Clearly clustered in the spectrum of contrasts in machine-learning analysis, taxonomic and functional family signatures can also be used for a diet and a life-style change advice.

Ratios of certain functional genomic characteristics (for instance, abundancy of genes belonging to aerobic versus anaerobic respiratory pathways, or of glycosyltransferases with different specificity) may be used as markers for fast screening of GM metagenomes with the following detailed taxonomic and functional analysis of cases where the `marker values seem to be alarming. This approach can become a part of a routine healthcare service, at least for the specific groups of patients. High abundancy of bacteriophages in a GM may be also considering as a reflection of a certain physiological stress [59], bacterial genome instability and increased virulence [60], and be used as an indicator of an individual's GM well-being. A number of medications that show a strong correlation with following shifts in GM structure (such as STATINS [57] or antipsychotics drugs [61]) are to be accompanied with a prescribed probiotic or anti-inflammatory treatment and diet recommendations.

Conclusion

By this small study on GMs of several individuals we show that GM metagenomics data can be used in preventive health care, but no clear translation of the correlations shown at a population level onto an individual's health can be easily performed. We can read individual GM metagenomes in details at taxonomic and functional levels and using cherry-picking approach we can suggest a number of plausible host's health problems, such as liver and heart problems, potential neural dysfunction, risk of gut inflammation and to distinguish GMs from the individuals treated by antibiotics. There is a level of uncertainty that we still need to deal with, especially in how we can integrate the quantitative GM communal information into a graded valuable health recommendations. We are working on algorithms for integration of the detectable information and are open for collaborations and more data to add to our database.

Conflicts of Interest

There are no conflicts of interest associated with this publication.

References

1. Ma N, Tian Y, Wu Y, Ma X (2017) Contributions of the Interaction between Dietary Protein and Gut Microbiota to Intestinal Health. Curr Protein Pept Sci 18: 795-808.
2. The integrative HMP (iHMP) research Network consortium (2019) The Integrative Human Microbiome Project. Nature 569: 641-648.
3. Qin Y, Wade PA (2018) Crosstalk between the microbiome and epigenome: messages from bugs. Biochem 163:105-112.
4. Thursby E, Juge N (2017) Introduction to the human gut microbiota. Biochem J 474: 1823-1836.
5. Rajilić-Stojanović M, de Vos WM (2014) The first 1000 cultured species of the human gastrointestinal microbiota. FEMS Microbiol Rev 38: 996-1047.
6. Hugon P, Dufour JC, Colson P, Fournier PE, Sallah K, et al. (2015) A comprehensive repertoire of prokaryotic species identified in human beings. Lancet Infect Dis 15: 1211-1219.
7. Human Microbiome Project Consortium (2012) Structure, function and diversity of the healthy human microbiome. Nature 486: 207-214.
8. Walker AW, Ince J, Duncan SH, Webster LM, Holtrop G, et al. (2011) Dominant and diet-responsive groups of bacteria within the human colonic microbiota. ISME J 5: 220-230.
9. Claesson MJ, Jeffery IB, Conde S, Power SE, O'Connor EM, et al. (2012) Gut microbiota composition correlates with diet and health in the elderly. Nature 488: 178-184.
10. Senghor B, Sokhna C, Ruimy R, Lagier JC (2018) Gut microbiota diversity according to dietary habits and geographical provenance. Hum Microbiome J 7-8: 1-9.
11. Zhernakova A, Kurilshikov A, Bonder MJ, Tigchelaar EF, Schirmer M, et al. (2016) Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. Science 352: 565-569.
12. Fortenberry JD (2013) The uses of race and ethnicity in human microbiome research. Trends Microbiol 21: 165-166.
13. Gupta VK, Paul S, Dutta C (2017) Geography, ethnicity or subsistence-specific variations in human microbiome composition and diversity. Front Microbiol 8:1162.
14. Falony G, Joossens M, Vieira-Silva S, Wang J, Darzi Y, et al. (2016) Population-level analysis of gut microbiome variation. Science 352:560-564.
15. Brooks AW, Sambhawa P, Blekhman R, Bordenstein SR (2018) Gut microbiota diversity across ethnicities in the United States. PLoS Biol 16: e2006842.
16. Arboleya S, Watkins C, Stanton C, Ross R (2016) Gut bifidobacteria populations in human health and aging. Front Microbiol 7: 1204.
17. Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, et al. (2012) Human gut microbiome viewed across age and geography. Nature 486: 222-227.
18. Goodrich JK, Waters JL, Poole AC, Sutter JL, Koren O, et al. (2014) Human genetics shape the gut microbiome. Cell 159: 789-799.

19. Schroeder BO, Backhed F (2016) Signals from the gut microbiota to distant organs in physiology and disease. Nat Med 22: 1079-1089.

20. Peterson DA, Frank DN, Pace NR, Gordon JI (2008) Metagenomic approaches for defining the pathogenesis of inflammatory bowel diseases. Cell Host Microbe 3: 417-427.

21. Sydney M (2011) *Desulfovibrio* species are potentially important in regressive autism. FinegoldMedical Hypotheses. 77: 270-274.

22. Midtvedt T (2012) The gut: a triggering place for autism – possibilities and challenges Microb Ecol Health Dis 23: 10.3402/mehd.v23i0.18982

23. Ding T, Schloss PD (2014) Dynamics and associations of microbial community types across the human body. Nature 509: 357-360.

24. Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, et al. (2011) Enterotypes of the human gut microbiome. Nature 473: 174-180

25. Jeffery IB, Claesson MJ, O'Toole PW, Shanahan F (2012) Categorization of the gut microbiota: enterotypes or gradients? Nat Rev Microbiol 10: 591-592.

26. Orakov AN, Sakenova NK, Sorokin A, Goryanin II (2017) ASAR: visual analysis of metagenomes in R. Bioinformatics 34: 1404-1405.

27. https://github.com/braithwaite/MGRASTer/

28. http://bioconductor.org/packages/release/bioc/html/metagenomeSeq.html

29. https://joey711.github.io/phyloseq/

30. Sonnenburg JL, Xu J, Leip DD, Chen CH, Westover BP, et al. (2005) Glycan foraging in vivo by an intestine-adapted bacterial symbiont. Science 307: 1955-1959.

31. La Rosa SL, Leth ML, Michalak L, Hansen ME, Pudlo NA, et al. (2019) The human gut Firmicute *Roseburia* intestinalis is a primary degrader of dietary β-mannans. Nature Comm 10: 905.

32. Hamer HM, Jonkers D, Venema K, Vanhoutvin S, Troost FJ, et al. (2008) Review article: the role of butyrate on colonic function. Aliment Pharmacol Ther 27: 104-119.

33. Roediger WE (1980) Role of anaerobic bacteria in the metabolic welfare of the colonic mucosa in man. Gut. 21: 793-798.

34. Wang HB, Wang PY, Wang X, Wan YL, Liu YC (2012) Butyrate enhances intestinal epithelial barrier function via up-regulation of tight junction protein Claudin-1 transcription. Dig Dis Sci 57: 3126-3135.

35. Wu GD, Chen J, Hoffmann C, Bittinger K, Chen YY, et al. (2011) Linking long-term dietary patterns with gut microbial enterotypes. Science 334: 105-108.

36. Moore EW, Holdeman LV (1974) Human fecal flora - normal flora of 20 Japanese-hawaiians. Appl Microbiol 27: 961-979.

37. Rausch P, Rehman A, Künzel S, Häsler R, Ott SJ, et al. (2011) Colonic mucosa-associated microbiota is influenced by an interaction of Crohn disease and FUT2 (Secretor) genotype. PNAS 108: 19030-19035.

38. Derrien M (2004) Akkermansia muciniphila gen. nov., sp. nov., a human intestinal mucin-degrading bacterium. Int J Syst Evol Microbiol 54: 1469-1476.

39. Li H, Limenitakis JP, Fuhrer T, Geuking MB, Lawson MA, et al. (2015) The outer mucus layer hosts a distinct intestinal microbial niche. Nat Commun 6: 8292

40. Winter SE, Winter MG, Xavier MN, Thiennimitr P, Poon V, et al. (2013) Host-derived nitrate boosts growth of *E. coli* in the inflamed gut. Science 339: 708-711.

41. Ridlon RJM, Kang DJ, Hylemon PB, Bajaj JS (2014) Bile acids and the gut microbiome. Curr Opin Gastroenterol 30: 332-338.

42. Kakiyama G, Pandak WM, Gillevet PM, Hylemon PB, Heuman DM, et al. (2013) Modulation of the fecal bile acid profile by gut microbiota in cirrhosis. J Hepatol 58: 949-955.

43. Dostal A, Chassard C, Hilty FM, Zimmermann MB, Jaeggi T, et al. (2011) Iron depletion and repletion with ferrous sulfate or electrolytic iron modifies the composition and metabolic activity of the gut microbiota in rats. J Nutr 142: 271-277.

44. Navarro-Garcia F, Gutierrez-Jimenez J, Garcia-Tovar C, Castro LA, Salazar-Gonzalez H, et al. (2010) Pic, an autotransporter protein secreted by different pathogens in the Enterobacteriaceae family, is a potent mucus secretagogue. Infect Immun 78: 4101-4109.

45. Devkota S, Chang EB (2015) Interactions between diet, bile acid metabolism, gut microbiota, and Inflammatory Bowel Diseases. Dig Dis 33: 351-356.

46. Davenport ER, Sanders JG, Song SJ, Amato KR, Clark AG, et al. (2017) The human microbiome in evolution. BMC Biol 15:120.

47. Morrison DJ, Preston T (2016) Formation of short chain fatty acids by the gut microbiota and their impact on human metabolism. Gut Microbes 7: 189-200.

48. Louis P, Flint HJ (2017) Formation of propionate and butyrate by the human colonic microbiota. Environ Microbiol 19: 29-41.

49. Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, et al. (2009) A core gut microbiome in obese and lean twins. Nature 457: 480-484.

50. Marcobal A, Barboza M, Sonnenburg ED, Pudlo N, Martens EC, et al. (2011) Bacteroides in the Infant Gut Consume Milk Oligosaccharides via Mucus-Utilization Pathways. Cell Host Microbe 10: 507-514.

51. Jakobsson HE, Rodríguez-Piñeiro AM, Schütte A, Ermund A, Boysen P, et al. (2015) The composition of the gut microbiota shapes the colon mucus barrier. EMBO Rep 16: 164-177.

52. Nakjang S, Ndeh DA, Wipat A, Bolam DN, Hirt RP (2012) A novel extracellular metallopeptidase domain shared by animal host-associated mutualistic and pathogenic microbes. PLoS ONE 7: e30287.

53. Luo Q, Kumar P, Vickers TJ, Sheikh A, Lewis WG, et al. (2014) Enterotoxigenic *Escherichia coli* secretes a highly conserved mucin-degrading metalloprotease to effectively engage intestinal epithelial cells. Infect Immun 82: 509-521.

54. Haussmann C, Rohdich F, Schmidt E, Bacher A, Richter G (1998) Biosynthesis of pteridines in *Escherichia coli* - Structural and mechanistic similarity of dihydroneopterin-triphosphate epimerase and dihydroneopterin aldolase. J Biol Chem 273: 17418-17424.

55. Stancu C, Sima A (2001) Statins: Mechanism of action and effects. J Cell Mol Med 5: 378-387.

56. Liu Y, Song X, Zhou H, Zhou X, Xia Y (2018) Gut microbiome associates with lipid-lowering effect of rosuvastatin *in vivo*. Front Microbiol. 9: 530.

57. Graziano TS, Cuzzullin MC, Franco GC, Schwartz-Filho HO, De Andrade ED, et al. (2015) Statins and antimicrobial effects: Simvastatin as a potential drug against *Staphylococcus aureus* biofilm. PLoS One 10: e0128098.

58. Rogers MAM, Aronoff DM (2016) The influence of non-steroidal anti-inflammatory drugs on the gut microbiome. Clin Microbiol Infect 22: 178.e1–178.e9.

59. Fang FC, Frawley ER, Tapscott T, Va´zquez-Torres A (2016) Bacterial stress responses during host infection. Cell Host Microbe 20:133-143.

60. Zuo T, Xiao-Juan L, Zhang Y, Cheung CP, Lam S, et al. (2019) Gut mucosal virome alterations in ulcerative colitis. Gut (BMJ) 68.

61. Bretler T, Weisberg H, Koren O, Neuman H (2019) The effects of antipsychotic medications on microbiome and weight gain in children and adolescents. BMC Med 17:112.