

Methane and fatty acid metabolism pathways are predictive of Low-FODMAP diet efficacy for patients with Irritable Bowel Syndrome

Ameen Eetemadi^{a,b} and Ilias Tagkopoulos^{a,b*}

^aDepartment of Computer Science, University of California, Davis, CA, USA; ^bGenome Center, University of California, Davis, CA, USA

*Correspondence: Ilias Tagkopoulos, itagkopoulos@ucdavis.edu

Methane and fatty acid metabolism pathways are predictive of Low-FODMAP diet efficacy for patients with irritable bowel syndrome

Objective. Identification of microbiota-based biomarkers as predictors of low-FODMAP diet response and design of a diet recommendation strategy for IBS patients.

Design. We created a compendium of gut microbiome and disease severity data before and after a low-FODMAP diet treatment from published studies followed by unified data processing, statistical analysis and predictive modeling. We employed data-driven methods that solely rely on the compendium data, as well as hypothesis-driven methods that focus on methane and short chain fatty acid (SCFA) metabolism pathways that were implicated in the disease etiology.

Results. The patient's response to a low-FODMAP diet was predictable using their pre-diet fecal samples with F1 accuracy scores of 0.750 and 0.875 achieved through data-driven and hypothesis-driven predictors, respectively. The fecal microbiome of patients with high response had higher abundance of methane and SCFA metabolism pathways compared to patients with no response (p -values $< 6 \times 10^{-3}$). The genera *Ruminococcus 1*, *Ruminococcaceae UCG-002* and *Anaerostipes* can be used as predictive biomarkers of diet response. Furthermore, the low-FODMAP diet followers were identifiable given their microbiome data (F1-score of 0.656).

Conclusion. Our integrated data analysis results argue that there are two types of patients, those with high colonic methane and SCFA production, who will respond well on a low-FODMAP diet, and all others, who would benefit a dietary supplementation containing butyrate and propionate, as well as probiotics with SCFA-producing bacteria, such as *Lactobacillus*. This work demonstrates how data integration can lead to novel discoveries and paves the way towards personalized diet recommendations for IBS.

Keywords: irritable bowel syndrome; low-FODMAP; diet; microbiome

Introduction

Irritable bowel syndrome (IBS) is a chronic gastrointestinal disorder that is prevalent in

approximately 11% of adult population.¹ It is associated with abdominal pain and changes in stool form and frequency of bowel movements.^{1,2} One of the emerging treatments for IBS is to reduce the amount of fermentable oligosaccharides, disaccharides, monosaccharides and polyols (FODMAPs) in the diet, also called the low-FODMAP diet, as recommended by the American College of Gastroenterology³ and the Canadian Association of Gastroenterology.⁴ The low-FODMAP diet has been effective for 50%-80% of IBS patients,⁵ however the patients who will benefit from this diet cannot be accurately identified beforehand. Several studies have attempted to create predictors for the efficacy of this diet in IBS using pre-treatment samples,⁶⁻⁸ however there is no evidence to show the utility of such a predictor across multiple studies. Furthermore, there is no common theory to explain the reason why the low-FODMAP diet is only effective for some patients in terms of disease etiology that is supported by data from multiple studies. It is believed that a low-FODMAP diet works by reducing the amount of carbohydrates that are not digested by the small intestine hence reach the colon to be used in gas producing microbial fermentation.⁹

Here, we investigate whether the efficacy of low-FODMAP diets on IBS patients can be predicted by analysis of easy to obtain biomarkers. Towards this goal, we created a compendium of microbiota metagenomics, by integrating data from 6 sources and fecal metagenomics samples from 152 unique IBS patients and 37 healthy adults. In addition, we investigated whether the amount of FODMAPs in an individual's diet, can be predicted using their gut microbiome data, showcasing the potential utility of microbiome data for assessing dietary adherence.

Materials and methods

Data curation

We searched PubMed for studies that have collected gut microbiome data before and after a period of low-FODMAP dietary treatment in humans. We found nine such studies and only six of them provided us with both the gut microbiome data as well as the corresponding metadata that is needed for this meta-analysis (**Table 1**). In all studies, the microbiome data came from fecal samples, characterized by 16S rRNA, or by the GA-map™ microbiome profiling.¹⁰ In GA-map™ microbiome profiling, each fecal sample is characterized by 54 numbers each representing the signal intensity of a DNA probe. The probes were designed for detection of bacterial taxonomies for distinguishing between IBS patients and healthy controls given fecal samples. The 16S rRNA and GA-map™ were analyzed independently.

Metadata processing

In all studies, the severity of IBS was quantified using IBS-SSS (IBS symptom severity scale) which is a number between zero and 500 representing the overall severity of IBS symptoms in a patient. We evaluated the patient's response to the diet based on the improvement in their IBS-SSS score ($\Delta_{\text{IBS-SSS}} = \text{IBS-SSS}_{\text{before}} - \text{IBS-SSS}_{\text{after}}$) after a period of following the low-FODMAP and labeled the patient's response as “High” (i.e. $\Delta_{\text{IBS-SSS}} \geq 150$), “Low” (i.e. $22 < \Delta_{\text{IBS-SSS}} < 150$), or “No” (i.e. $\Delta_{\text{IBS-SSS}} \leq 22$). The high threshold of 150 is reasonable since the reported mean plus standard deviation of $\Delta_{\text{IBS-SSS}}$ for a placebo treatment can range from 124 to 162,^{11,12} and therefore a “High” response is unlikely to be associated with a placebo effect. The low threshold of 22 was chosen to create a balance between the “No” and “High” response groups.

Preprocessing of 16S rRNA microbiome data

We analyzed 16S rRNA data separately for each study before integration. We used DADA2¹³ version 1.10.1 implemented in R version 3.5.2 following the package's online tutorial (benjjneb.github.io/dada2/bigdata.html). First, primer and adapter sequences were removed from each read and quality control was performed by removing 16S rRNA reads that were chimeric, shorter than 260 bp, or had at least two expected errors. In addition, longer reads were truncated at 260 bp since read qualities decreased sharply afterward. For one dataset¹⁴, the reverse reads were truncated at 160 bp instead due to the decrease of read qualities at lower base pairs compared to the forward reads. Next, we performed de novo sequence assembly to identify operational taxonomic units (OTUs). Then SILVA database¹⁵ version 32 was used to identify bacterial taxonomies associated with 16S rRNA assembled sequences. Taxa that were only observed in a single sample were filtered out.

Functional profiling from 16S rRNA microbiome data

We imported OTU read counts of the DADA2 analysis into qiime2, searched against Greengenes¹⁶ and filtered out OTUs that could not be matched at the 97% identity threshold as needed for PICRUSt.¹⁷ Samples with no remaining OTUs were removed if any, and predictive metagenome profiling and KEGG pathway enrichment analysis (for level L3) were performed using PICRUSt. Finally, we converted the read counts to relative abundances and transformed using centered log-ratio transform (CLR) to account for the compositionality of microbiome data.¹⁸ In the case of zero relative abundances of a given pathway, we used the minimum amongst CLR transformed values of non-zero read counts, subtracted by 10% of their standard deviations. Given that reported KEGG pathways from PICRUSt did not include specific pathway for

SCFAs, we relied on fatty-acid pathway abundances instead.

GA-map™ microbiome data processing

We normalized the signal intensities of 54 probes from each study separately to have zero-mean and unit-variance for a given probe before integration. To estimate the relative enrichment of methane metabolism in gut microbiome, we used the AG0581 probe (designed for detection of genus Dorea). The genus Dorea has been shown previously to be negatively associated with breath methane levels (See¹⁹, Table 3). To estimate the enrichment of SCFA metabolism pathways in gut microbiome, we used two pairs of probes AG0686, AG1099 (designed for genus Parabacteroides) and AG1225, AG1226 (designed for genus Alistipes) as their corresponding genus have been shown to be negatively associated with fecal SCFA levels (See²⁰, Table S5).

Differential analysis and statistical validation

We used unpaired non-parametric Wilcoxon rank-sum test for identifying pathways and taxa that are differentially abundant between IBS patients with high (n=8), low (n=29), or no (n=9) response to low-FODMAP diet where degrees of freedom is equal to the number of samples used minus two (e.g. degrees of freedom for high versus no response was 8+9-2=15). The calculated p-values were one sided for hypothesis-driven statistical validations and two sided for data-driven differential analysis. We also calculate FDR-corrected p-values (i.e. q-values) in data-driven differential analysis to account for multiple hypothesis testing given the number of KEGG pathways (n=237) and genus taxa used (n=217), with thresholds of 0.15 or lower.

Diet response prediction

We first integrated data from multiple studies and performed dimensionality reduction using sparse principal component analysis^{21,22} reducing the number of microbiome features (microbial taxa, enriched pathways or GA-map probes) to 30% of the number of profiles in the dataset. Then for a given pair of classification labels, we created random forest (RF) classifier and evaluated using leave-one-out cross-validation. We also evaluated the classification performance by iterative removal of the feature that is identified as least important by the RF classifier until only one feature remained. In all cases the areas under the precision-recall (PR) and receiver operating characteristic (ROC) curves, as well as the F1 score (the harmonic mean of precision and recall) were calculated.

Results

Figure 1 illustrates our data analysis methodology. A consistent data processing pipeline was applied to the curated metagenomics data enabling downstream analysis (hypothesis-driven and data-driven). The hypothesis-driven analysis was informed from the illustrated literature-based hypotheses: (a) the methane gas can inhibit intestinal motility hence contributing to stool abnormality in the form of constipation or bloating,²³ (b) methanogenesis requires hydrogen and carbon dioxide that can be generated by anaerobic fermentation of undigested carbohydrates in colon,²⁴ and (c) short-chain fatty acids (SCFAs) such as formate can also induce methanogenesis independently or in tandem with hydrogen.^{9,25} Therefore, in hypothesis-driven analysis we only used methane and fatty acid metabolism pathway abundances as input while in data-driven analysis all pathways and taxa (at genus level) were used for differential abundance analysis and predictive modelling.

Comparison of high/low response to Low-FODMAP diet reveals structural differences in the microbiota

Pre-diet fecal metagenomes of IBS patients were integrated and processed from five studies along with disease severity scores (IBS-SSS) ranging from zero to 500 before and after following a low-FODMAP for a total of 152 patients (**Figure 2A**). For differential analysis, we focused on the patients with most extreme responses (high versus no response) that had 16S rRNA metagenomic profiles (n=17). Top 5 KEGG pathways were differentially abundant with q-values < 0.11 with fatty acid metabolism being the most differentially expressed. However, there was no differentially abundant genus taxa when a q-value significance threshold of 0.15 is used (**Figure 2. B-C**). Three genera (*Ruminococcaceae* UCG-002, *Ruminococcus* 1 and *Anaerostipes*) were identified amongst the top 5 to be positively associated with stool SCFA levels based on other studies.^{20,26} Therefore a 3-genus microbiome biomarker was designed by adding their CLR-transformed abundances providing higher values for patients with a high response versus low response (p-value = 1.0×10^{-10}) or no response (p-value = 2.5×10^{-4}) following the diet (**Supplementary Figure S5.**) Note that the microbiome profiles of patients with low response were never used in the discovery of top five genera reported in **Figure 2. C**. A data-driven predictor of high/no response was built given all KEGG pathway abundances providing an F1 score of 0.750, AUROC of 0.708 (baseline: 0.5) and AUPR of 0.629 (baseline: 0.471).

We also created predictor for high versus low or no response for patients with 16S rRNA metagenome profiles (**Figure 3. A, B, D and D**). Using pathway abundances as input provides an F1 score of 0.625, AUROC of 0.850 (baseline: 0.5) and AUPR of 0.693 (baseline: 0.174) while with genus taxa abundances as input an F1 score of 0.533, AUROC of 0.873 (baseline: 0.5) and AUPR of 0.425 (baseline: 0.174) was achieved.

For patients with GA-map data (**Figure 3. C and F**) an F1 score of 0.581, AUROC of 0.625 (baseline: 0.5) and AUPR of 0.530 (baseline: 0.462) was achieved.

IBS patients with methanogenic fecal microbiome respond better to Low-FODMAP diets

Low intestinal motility of IBS patients has been associated with intestinal production of methane²³ due to methane producing microbes (methanogens) in the gut,^{20,27} which use undigested carbohydrates for their metabolism.⁹ Therefore, we hypothesized that response to low-FODMAP diet is associated with gut microbiome methane metabolism capability. To validate this hypothesis, we performed meta-analysis on 46 patients; integrated from three studies^{14,28,29} that rely on 16S rRNA data. In agreement with our hypothesis, the high response group of patients had a significantly higher enrichment in methane metabolism pathway of their pre-treatment microbiome samples compared to low response (p-value = 1.3×10^{-2}) and no response (p-value = 5.6×10^{-3}) groups (**Figure 4A**). We then used GA-map microbiome data from a separate study⁶ with 31 IBS patients, using only the probe associated with methane production. The analysis of GA-map data also supports our hypothesis with high response patients having higher abundance in methane production associated taxa when compared to the no response patients (p-value = 7.4×10^{-3}) (**Supplementary Figure S1**).

The efficacy of Low-FODMAP diet can be accurately predicted by methane and short-chain fatty acid metabolic pathways

Short-chain fatty acids (SCFAs) are key products of microbial fermentation in human intestine and important for health of epithelial cells.³⁰ Therefore, we also analyzed the enrichment of fatty-acid metabolism pathway in 16S rRNA fecal microbiome data of

IBS patients. Our analysis shows higher enrichment in fatty-acid metabolism for high versus no response patients ($p\text{-value} = 7.8 \times 10^{-4}$) (**Figure 4B**). Next, we created a classifier to predict the patient's response (high versus no response) based on methane and fatty acid metabolism in 16s rRNA data. Our random forest (RF) classifier achieved 0.89 and 0.84 for area under the curve (AUC) of ROC and PR curves, respectively (**Figure 4C-D**). We also performed analysis for GA-map probe data using taxa probes that have been associated with SCFA levels in fecal samples, but did not find a significant difference between the "High-response" and "Low-response" IBS patients.

Predicting diets from their effect on the microbiome

Diet is considered to be an important factor for modulating intestinal microbiota,³¹ however it is not clear whether a low-FODMAP diet leads into common changes in gut microbiome across different individuals. To investigate this, we used 188 16S rRNA fecal microbiome profiles from IBS patients and healthy individuals before ($n=95$) and after ($n=93$) low-FODMAP dietary intervention. Microbiome samples were characterized by their KEGG pathway and genus taxa abundances. We used random forest classifier to predict whether the microbiome sample is taken before, or after the low-FODMAP dietary intervention (**Figure 5. A**). When pathway abundances were used as input the classifier achieved F1 score of 0.656, AUROC of 0.687 (baseline: 0.5) and AUPR of 0.663 (baseline: 0.495) (**Figure 5 B**). Only three pathways were needed to achieve an F1 score of 0.66 (**Figure 5 D**). Using taxa abundances at genus level for classification provided F1 score of 0.602, AUROC of 0.608 (baseline: 0.5) and AUPR of 0.597 (baseline: 0.495) (**Figure 5. C**).

Discussion

While several studies have confirmed the efficacy of low-FODMAP diet for symptom management in IBS, between 55%-66% of IBS patients have a response that is similar to a placebo treatment (**Supplementary Figure S4.**). We hypothesized that the patient's response level (high/low/no) to a low-FODMAP diet can be predicted using their fecal microbiome samples. Although this hypothesis had been validated to an extent by individual studies, there is no predictor that (a) works across multiple studies and (b) comes with a mechanistic explanation of the patient's response based on their microbiomes. To this end we integrated data from five distinct studies and performed a meta-analysis showing that the patient's response to low-FODMAP diet is predictable given their fecal microbiome. We also formed a literature-based hypothesis supported by the integrated data that a high response to low-FODMAP diet is associated with higher abundance of methane and SCFA metabolism pathways in gut microbiome. Our mechanistic explanation is that a low-FODMAP diet works by lowering the amount of colonic methane that is shown to slow down intestinal motility,²³ a precursor to constipation and/or bloating. Therefore, patients with highest response have a colonic microbiome with substantial methane production capability due to (a) methane metabolism pathways, and (b) SCFA metabolism pathways that promote methanogenesis, both of which rely on microbial digestion of carbohydrates. Gut microbes can also use formate or hydrogen to produce acetate,³² an SCFA with anti-inflammatory properties,³³ which may inhibit their availability for methanogenesis and decrease bloating. The microbiome SCFA pathways can have positive or negative impact on microbial secretion and absorption of gases, which necessitates more in-depth investigation of their role in IBS dietary treatments (e.g. low-FODMAP diet and probiotics). Additionally, we showed that gut microbiome data can be used to predict whether a patient is following a low-FODMAP diet, suggesting that this diet modulates gut microbiome and leaves

identifiable traces which can be used for assessing dietary compliance. This work showcases the utility of integrated meta-analysis using raw data from individual studies with a consistent methodology to arrive at new insights (see **Supplementary Figure S6.**). Although there were several differences amongst the low-FODMAP studies that can create risks for data analysis (see the “Differences amongst studies” tab under **Supplementary Data.xlsx**), we found no significant change in the amount of improvement of IBS-SSS score after following a low-FODMAP diet amongst the studies despite their differences (see **Supplementary Figure S9.**). In addition, when it comes to microbiome data processing and analysis, we minimized the impact of such differences by applying the same standard pipeline starting from the raw microbiome data of each study. We acknowledge that the other differences (e.g. stool sample handling and metagenomic sequencing) can also be problematic in revealing any signal, however once such pattern is discovered, these differences increase the robustness and reproducibility of the analysis, as it becomes less sensitive to the specific details of the techniques used.

Prior studies show that lower abundance of microorganisms that produce butyrate (an important SCFA) is associated with irritable bowel syndrome,³⁴ *Lactobacillus* based probiotics promote production of SCFAs in the gut³⁵ and improve disease symptoms in IBS.³⁶ Consistent with our meta-analysis results, we suggest a biomarker-based diet recommendation system where a low-FODMAP diet is recommended to patients with high colonic methane and SCFA production, and a probiotic supplementation with SCFA producing microbes is recommended to patients with low colonic methane and SCFA production. Such a personalized recommendation system will be inline with dietary recommendations from the American College of Gastroenterology and the Canadian Association of Gastroenterology for IBS which consider both dietary treatments as beneficial,^{3,4} while expected to decrease the array of

treatments that patients need to try before finding the treatment that works for them. Clinical trials will be necessary to identify best biomarkers, probiotic species and dosages and evaluate the patient's response compared to alternative treatments. A comprehensive array of tests including gas analysis of breath samples, shotgun metagenomics, qPCR with primers that can detect SCFA producing microbiomes and methanogenic microorganisms that are archaeal, and gas chromatography–mass spectrometry (GC/MS) for detecting SCFA levels from microbiome samples (fecal or through colonic biopsy), will be necessary to provide more accurate insight into the microbiome pathways discussed. Given the advent of low-cost breath testing and accessibility of primer-based qPCR testing of fecal samples, gut microbiome methane and SCFA metabolism levels can be readily assessed in the clinic in order to provide more effective dietary recommendations for IBS patients. Intestinal bacterial infections are commonly diagnosed through low-cost qPCR testing of stool samples for detection of known pathogens given target-specific primers.³⁷ Intestinal malabsorption of carbohydrates is also diagnosed in the clinic using hydrogen and methane breath testing although with variable repeatability.³⁸ Upon development of a qPCR kit for gut microbiome SCFA metabolism estimation (e.g. by detection of *Ruminococcus 1*, *Ruminococcaceae* UCG-002 and *Anaerostipes* genera levels), a personalized IBS diet can be employed in the clinic where SCFA supplementation (prebiotic or postbiotic) is recommended when SCFA microbiome metabolism is low, and a low-FODMAP diet is recommended when SCFA and methane metabolism of the gut microbiome are above a calibrated threshold. We believe that the recent advances in high resolution omics and computational methods across diet, microbiome, and health,³⁹ as well as novel ways of food representation that rely on artificial intelligence,^{40,41} will give rise to more personalized dietary treatments potentially revolutionizing clinical nutrition.

It is important to note that, the analyzed data here included microbiome profiles from IBS patients with diarrhea, constipation, or both symptoms, however, we did not perform a separate analysis based on the IBS type since multiple studies did not provide the IBS type information per patient. Further studies will be necessary to validate the hypothesized mode of action for this diet in reducing constipation and bloating symptoms of IBS, and to understand the possibly different modes of action in reducing diarrhea.

Acknowledgements

This work was supported by the USDA/NSF AI Institute for Next Generation Food Systems (AIFS), USDA award number 2020-67021-32855 and NSF award 1743101 to I.T.

Declaration of interest statement

No potential competing interest was reported by the authors.

References

1. Lovell RM, Ford AC. Global prevalence of and risk factors for irritable bowel syndrome: a meta-analysis. *Clinical gastroenterology and hepatology*. 2012;10(7):712–721.
2. Chey WD, Kurlander J, Eswaran S. Irritable bowel syndrome: a clinical review. *Jama*. 2015;313(9):949–958.
3. Ford AC, Moayyedi P, Chey WD, Harris LA, Lacy BE, Saito YA, Quigley EM, others. American College of Gastroenterology monograph on management of irritable bowel syndrome. *American Journal of Gastroenterology*. 2018;113:1–18.
4. Moayyedi P, Andrews CN, MacQueen G, Korownyk C, Marsiglio M, Graff L, Kvern B, Lazarescu A, Liu L, Paterson WG, et al. Canadian Association of Gastroenterology clinical practice guideline for the management of irritable bowel syndrome (IBS). *Journal of the Canadian Association of Gastroenterology*. 2019;2(1):6–29.
5. Staudacher HM, Whelan K. The low FODMAP diet: recent advances in understanding its mechanisms and efficacy in IBS. *Gut*. 2017;66(8):1517–1527.

6. Bennet SM, Böhn L, Störsrud S, Liljebo T, Collin L, Lindfors P, Törnblom H, Öhman L, Simrén M. Multivariate modelling of faecal bacterial profiles of patients with IBS predicts responsiveness to a diet low in FODMAPs. *Gut*. 2018;67(5):872–881.
7. Rossi M, Aggio R, Staudacher HM, Lomer MC, Lindsay JO, Irving P, Probert C, Whelan K. Volatile organic compounds in feces associate with response to dietary intervention in patients with irritable bowel syndrome. *Clinical Gastroenterology and Hepatology*. 2018;16(3):385–391.
8. Wilder-Smith C, Olesen SS, Materna A, Drewes A. Predictors of response to a low-FODMAP diet in patients with functional gastrointestinal disorders and lactose or fructose intolerance. *Alimentary pharmacology & therapeutics*. 2017;45(8):1094–1106.
9. Kalantar-Zadeh K, Berean KJ, Burgell RE, Muir JG, Gibson PR. Intestinal gases: influence on gut disorders and the role of dietary manipulations. *Nature Reviews Gastroenterology & Hepatology*. 2019:1–15.
10. Casen C, Vebø H, Sekelja M, Hegge F, Karlsson M, Cierniejewska E, Dzankovic S, Frøyland C, Nestestog R, Engstrand L, et al. Deviations in human gut microbiota: a novel diagnostic test for determining dysbiosis in patients with IBS or IBD. *Alimentary pharmacology & therapeutics*. 2015;42(1):71–83.
11. Lyra A, Hillilä M, Huttunen T, Männikkö S, Taalikka M, Tennilä J, Tarpila A, Lahtinen S, Ouwehand AC, Veijola L. Irritable bowel syndrome symptom severity improves equally with probiotic and placebo. *World journal of gastroenterology*. 2016;22(48):10631.
12. Kaptchuk TJ, Friedlander E, Kelley JM, Sanchez MN, Kokkotou E, Singer JP, Kowalczykowski M, Miller FG, Kirsch I, Lembo AJ. Placebos without deception: a randomized controlled trial in irritable bowel syndrome. *PloS one*. 2010;5(12):e15591.
13. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: high-resolution sample inference from Illumina amplicon data. *Nature methods*. 2016;13(7):581–583.
14. McIntosh K, Reed DE, Schneider T, Dang F, Keshteli AH, De Palma G, Madsen K, Bercik P, Vanner S. FODMAPs alter symptoms and the metabolome of patients with IBS: a randomised controlled trial. *Gut*. 2017;66(7):1241–1251.
15. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic acids research*. 2012;41(D1):D590–D596.
16. McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A, Andersen GL, Knight R, Hugenholtz P. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *The ISME journal*. 2012;6(3):610.
17. Langille MG, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA, Clemente JC, Burkepille DE, Thurber RLV, Knight R, et al. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nature biotechnology*. 2013;31(9):814.

18. Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. Microbiome datasets are compositional: and this is not optional. *Frontiers in microbiology*. 2017;8:2224.
19. Parthasarathy G, Chen J, Chen X, Chia N, O'Connor HM, Wolf PG, Gaskins HR, Bharucha AE. Relationship between microbiota of the colonic mucosa vs feces and symptoms, colonic transit, and methane production in female patients with chronic constipation. *Gastroenterology*. 2016;150(2):367–379.
20. Sloan TJ, Jalanka J, Major GA, Krishnasamy S, Pritchard S, Abdelrazig S, Korpela K, Singh G, Mulvenna C, Hoad CL, et al. A low FODMAP diet is associated with changes in the microbiota and reduction in breath hydrogen but not colonic volume in healthy subjects. *PloS one*. 2018;13(7):e0201410.
21. Sigg C. nsprcomp v0. 5.1-2.
22. Sigg CD, Buhmann JM. Expectation-maximization for sparse and non-negative PCA. In: *Proceedings of the 25th international conference on Machine learning*. 2008. p. 960–967.
23. Triantafyllou K, Chang C, Pimentel M. Methanogens, methane and gastrointestinal motility. *Journal of neurogastroenterology and motility*. 2014;20(1):31.
24. Levitt MD. Production and excretion of hydrogen gas in man. *New England Journal of Medicine*. 1969;281(3):122–127.
25. Thiele JH, Zeikus JG. Control of interspecies electron flow during anaerobic digestion: significance of formate transfer versus hydrogen transfer during syntrophic methanogenesis in flocs. *Applied and environmental Microbiology*. 1988;54(1):20–29.
26. Yamamura R, Nakamura K, Kitada N, Aizawa T, Shimizu Y, Nakamura K, Ayabe T, Kimura T, Tamakoshi A. Associations of gut microbiota, dietary intake, and serum short-chain fatty acids with fecal short-chain fatty acids. *Bioscience of microbiota, food and health*. 2019:19–010.
27. Kim G, Deepinder F, Morales W, Hwang L, Weitsman S, Chang C, Gunsalus R, Pimentel M. *Methanobrevibacter smithii* is the predominant methanogen in patients with constipation-predominant IBS and methane on breath. *Digestive diseases and sciences*. 2012;57(12):3213–3218.
28. Harvie RM, Chisholm AW, Bisanz JE, Burton JP, Herbison P, Schultz K, Schultz M. Long-term irritable bowel syndrome symptom control with reintroduction of selected FODMAPs. *World journal of gastroenterology*. 2017;23(25):4632.
29. Schumann D, Langhorst J, Dobos G, Cramer H. Randomised clinical trial: yoga vs a low-FODMAP diet in patients with irritable bowel syndrome. *Alimentary pharmacology & therapeutics*. 2018;47(2):203–211.
30. Gill P, Van Zelm M, Muir J, Gibson P. Short chain fatty acids as potential therapeutic agents in human gastrointestinal and inflammatory disorders. *Alimentary pharmacology & therapeutics*. 2018;48(1):15–34.

31. Kolodziejczyk AA, Zheng D, Elinav E. Diet–microbiota interactions and personalized nutrition. *Nature Reviews Microbiology*. 2019;1–12.
32. Smith NW, Shorten PR, Altermann EH, Roy NC, McNabb WC. Hydrogen cross-feeders of the human gastrointestinal tract. *Gut Microbes*. 2019;10(3):270–288.
33. Maslowski KM, Vieira AT, Ng A, Kranich J, Sierro F, Yu D, Schilter HC, Rolph MS, Mackay F, Artis D, et al. Regulation of inflammatory responses by gut microbiota and chemoattractant receptor GPR43. *Nature*. 2009;461(7268):1282–1286.
34. Pozuelo M, Panda S, Santiago A, Mendez S, Accarino A, Santos J, Guarner F, Azpiroz F, Manichanh C. Reduction of butyrate-and methane-producing microorganisms in patients with Irritable Bowel Syndrome. *Scientific reports*. 2015;5:12693.
35. Nagpal R, Wang S, Ahmadi S, Hayes J, Gagliano J, Subashchandrabose S, Kitzman DW, Becton T, Read R, Yadav H. Human-origin probiotic cocktail increases short-chain fatty acid production via modulation of mice and human gut microbiome. *Scientific reports*. 2018;8(1):1–15.
36. Pedersen N, Andersen NN, Végh Z, Jensen L, Ankersen DV, Felding M, Simonsen MH, Burisch J, Munkholm P. Ehealth: low FODMAP diet vs *Lactobacillus rhamnosus* GG in irritable bowel syndrome. *World Journal of Gastroenterology: WJG*. 2014;20(43):16215.
37. Momčilović S, Cantacessi C, Arsić-Arsenijević V, Otranto D, Tasić-Otašević S. Rapid diagnosis of parasitic diseases: current scenario and future needs. *Clinical Microbiology and Infection*. 2019;25(3):290–309.
38. Yao CK, Tuck CJ. The clinical value of breath hydrogen testing. *Journal of Gastroenterology and Hepatology*. 2017;32:20–22.
39. Eetemadi A, Rai N, Pereira BMP, Kim M, Schmitz H, Tagkopoulos I. The Computational Diet: A Review of Computational Methods Across Diet, Microbiome, and Health. *Frontiers in Microbiology*. 2020;11:393. doi:10.3389/fmicb.2020.00393
40. Dooley DM, Griffiths EJ, Gosal GS, Buttigieg PL, Hoehndorf R, Lange MC, Schriml LM, Brinkman FS, Hsiao WW. FoodOn: a harmonized food ontology to increase global food traceability, quality control and data integration. *npj Science of Food*. 2018;2(1):1–10.
41. Youn J, Naravane T, Tagkopoulos I. Using Word Embeddings to Learn a Better Food Ontology. *Frontiers in Artificial Intelligence*. 2020;3:93. doi:10.3389/frai.2020.584784
42. Chumpitazi BP, Hollister EB, Oezguen N, Tsai CM, McMeans AR, Luna RA, Savidge TC, Versalovic J, Shulman RJ. Gut microbiota influences low fermentable substrate diet efficacy in children with irritable bowel syndrome. *Gut microbes*. 2014;5(2):165–175.
43. Chumpitazi BP, Cope JL, Hollister EB, Tsai CM, McMeans AR, Luna RA, Versalovic J, Shulman RJ. Randomised clinical trial: gut microbiome biomarkers are

associated with clinical response to a low FODMAP diet in children with the irritable bowel syndrome. *Alimentary pharmacology & therapeutics*. 2015;42(4):418–427.

44. Staudacher HM, Lomer MC, Farquharson FM, Louis P, Fava F, Franciosi E, Scholz M, Tuohy KM, Lindsay JO, Irving PM, et al. A diet low in FODMAPs reduces symptoms in patients with irritable bowel syndrome and a probiotic restores bifidobacterium species: a randomized controlled trial. *Gastroenterology*. 2017;153(4):936–947.

45. Valeur J, Småastuen MC, Knudsen T, Lied GA, Røseth AG. Exploring gut microbiota composition as an indicator of clinical response to dietary FODMAP restriction in patients with irritable bowel syndrome. *Digestive Diseases and Sciences*. 2018;63(2):429–436.

Tables

Table 1. Studies with gut microbiome data involving low-FODMAP dietary treatment.

Id	Reference	Microbiome Technology	Access
1	42	16s rRNA	N/A
2	43	16s rRNA	N/A
3	28	16s rRNA	Granted
4	14	16s rRNA	Granted
5	44	16s rRNA	N/A
6	6	GA-map™	Granted
7	20	16s rRNA	Granted
8	29	16s rRNA	Granted
9	45	GA-map™	Granted

N/A: Authors did not grant access to metadata and/or raw microbiome data.

Figure captions

Figure 1. Overview of low-FODMAP diet response prediction for irritable bowel syndrome (IBS): The response of IBS patients to a low-FODMAP diet and their pre-diet fecal metagenomes were integrated and analyzed from five independent studies. Consistent data processing pipeline was applied on raw metagenome data to infer the relative pathway and taxa (at genus level) abundances for individual gut microbiomes. In a data-driven analysis, differentially abundant taxa and pathways were identified for patients with high versus no response to the low-FODMAP die. Diet response predictors were built to identify whether an IBS patient will benefit from a low-FODMAP diet given their pre-diet fecal metagenome. Furthermore a hypothesis-driven analysis was performed given the hypothesized relationships between FODMAPs, methane metabolism, fatty acid metabolism and illustrated colon functions base on literature. Although similar to the data-driven analysis, only the pathway abundances relating to methane and fatty acid metabolism were used for statistical validation, model training and the final diet response predictor.

Figure 2. Pre-diet microbial differential abundances for IBS patients with high versus no response to the low-FODMAP diet: (A) IBS patient records from five studies are sorted into three groups based on their response to the low-FODMAP diet (High/Low/No) given the amount of improvement in IBS symptom severity after following the diet. (B) Top 5 pre-diet gut microbiome KEGG pathways that are differentially abundant (following a clr-transformation of their relative abundances) amongst High versus No response patient groups (q-values < 0.11; Fatty acid metabolism p-value = 1.5×10^{-3} ; Nucleotide excision repair p-value = 3.7×10^{-3} ; Phenylalanine, tyrosine and tryptophan biosynthesis p-value = 3.7×10^{-3} ; RNA polymerase p-value = 3.7×10^{-3} ; Thiamine metabolism p-value = 3.7×10^{-3}). (C) Similar to (B) for differentially abundant genus

taxa (genus related q-values are not significant using a threshold of 0.15;
Ruminococcaceae UCG-002 p-value = 3.1×10^{-3} ; Ruminococcus 1 p-value = 1.3×10^{-2} ;
Victivallis p-value = 2.3×10^{-2} ; Anaerostipes p-value = 3.0×10^{-2} ; Turicibacter p-value =
 6.0×10^{-2}).

Figure 3. Prediction of response to low-FODMAP diet given pre-diet microbiome data: (A-C) ROC and PR curves for prediction of response to low-FODMAP diet using pathway abundances, genus taxa abundances and GA-map probe signals of pre-diet gut microbiome. The star relates to the threshold used for calculating the F1 scores. (D-F) The F1 scores relating to predictive models when the least important feature (pathway, taxa or GA-map probe) is incrementally removed until only a single feature remains in the predictive model. The stars highlight the best F1 score achieved and each corresponds to a pair of ROC and PR pair curves on the top (i.e. A&D, B&E and C&F correspond respectively).

Figure 4. Prediction of response to low-FODMAP diet given pre-diet microbial abundances for methane and fatty acid metabolism pathways. (A&B) Methane and fatty acid metabolism pathway enrichment of pre-treatment gut microbiome for patients with High, Low or No response to low-FODMAP diet. (C&D) ROC and PR curves for predicting High vs. No response to low-FODMAP diet using methane and fatty acid metabolism pathway abundances (CLR-transformed) in gut microbiome. (E&F) ROC and PR curves for predicting High vs. Low or No response to low-FODMAP diet using methane and fatty acid metabolism pathway abundances (CLR-transformed) in gut microbiome.

Figure 5. Prediction of diet (low-FODMAP vs. other) given microbiome data: (A) Fecal metagenomes were integrated from four studies along with the dietary regimen that was

followed prior to sampling. Consistent data processing pipeline was applied on raw metagenome data to infer the relative pathway and taxa (at genus level) abundances for each sample. Diet predictors were built to identify the individual's diet given their fecal metagenome. (B & C) ROC and PR curves for diet prediction using pathway and genus taxa abundances in gut microbiome. The star relates to the threshold used for calculating the F1 scores. (D-E) The F1 scores relating to predictive models when the least important feature (pathway or taxa) is incrementally removed until only a single feature remains in the predictive model. The stars highlight the best F1 score achieved and each corresponds to a pair of ROC and PR pair curves on the top (i.e. B&D and C&E correspond respectively).