Check for updates

RESEARCH NOTE

REVISED Investigation of gut microbiome association with inflammatory bowel disease and depression: a machine learning approach [version 2; peer review: 2 approved with reservations]

Pedro Morell Miranda , Francesca Bertolini, Haja N. Kadarmideen

Department of Bio and Health Informatics, Technical University of Denmark, Kongens Lyngby, 2800, Denmark

V2 First published: 05 Jun 2018, 7:702
<https://doi.org/10.12688/f1000research.15091.1>
 Latest published: 17 Apr 2019, 7:702
<https://doi.org/10.12688/f1000research.15091.2>

Abstract

Background: Inflammatory bowel disease (IBD) is a group of chronic diseases related to inflammatory processes in the digestive tract generally associated with an immune response to an altered gut microbiome in genetically predisposed subjects. For years, both researchers and clinicians have been reporting increased rates of anxiety and depression disorders in IBD, and these disorders have also been linked to an altered microbiome. However, the underlying pathophysiological mechanisms of comorbidity are poorly understood at the gut microbiome level.

Methods: Metagenomic and metatranscriptomic data were retrieved from the Inflammatory Bowel Disease Multi-Omics Database. Samples from 70 individuals that had answered to a self-reported depression and anxiety questionnaire were selected and classified by their IBD diagnosis and their questionnaire results, creating six different groups. The cross-validation random forest algorithm was used in 90% of the individuals (training set) to retain the most important species involved in discriminating the samples without losing predictive power. The validation set that represented the remaining 10% of the samples equally distributed across the six groups was used to train a random forest using only the species selected in order to evaluate their predictive power.

Results: A total of 24 species were identified as the most informative in discriminating the 6 groups. Several of these species were frequently described in dysbiosis cases, such as species from the genus *Bacteroides* and *Faecalibacterium prausnitzii*. Despite the different compositions among the groups, no common patterns were found between samples classified as depressed. However, distinct taxonomic profiles within patients of IBD depending on their depression status were detected.

Conclusions: The machine learning approach is a promising approach

Open Peer Review

Approval Status

	1	2
version 2		
(revision)		
17 Apr 2019		
version 1		
05 Jun 2018		

1. **Yasir Suhail** , Yale University, Storrs, USA
2. **Manikandan Narayanan**, Indian Institute of Technology (IIT) Madras, Chennai, India

Any reports and responses or comments on the article can be found at the end of the article.

for investigating the role of microbiome in IBD and depression. Abundance and functional changes in these species suggest that depression should be considered as a factor in future research on IBD.

Keywords

Inflammatory Bowel Disease, Depression, Microbiome, Machine Learning, Random Forest, Metagenomic, Metatranscriptomic.

This article is included in the **Bioinformatics** gateway.

Corresponding author: Haja N. Kadarmideen (hajak@dtu.dk)

Author roles: **Morell Miranda P:** Conceptualization, Formal Analysis, Investigation, Writing – Original Draft Preparation; **Bertolini F:** Supervision, Writing – Review & Editing; **Kadarmideen HN:** Funding Acquisition, Project Administration, Supervision, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: Authors thank funding from The Danish Independent Research Council (Technology and Production) grant number DFF – 6111-00471. P.M.M. thanks funding from Technical University of Denmark, DTU Bioinformatics for his research assistantship to conduct this study.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2019 Morell Miranda P *et al.* This is an open access article distributed under the terms of the **Creative Commons Attribution License**, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Morell Miranda P, Bertolini F and Kadarmideen HN. **Investigation of gut microbiome association with inflammatory bowel disease and depression: a machine learning approach [version 2; peer review: 2 approved with reservations]** F1000Research 2019, 7:702 <https://doi.org/10.12688/f1000research.15091.2>

First published: 05 Jun 2018, 7:702 <https://doi.org/10.12688/f1000research.15091.1>

REVISED Amendments from Version 1

Main difference from previous version and the version 2 is that as per Reviewer 1's comments, we have added a more thorough description of the dataset and a workflow illustrating the k-Fold Cross Validation approach for Random Forest (new Figure 1). As per the comments of Reviewer 2, we have rerun the species identification considering all IBD samples as a group and then classified by a depressed or not depressed state. We have furthermore added a small description of the set of species and expanded the introduction with some of the citations suggested. However, we cannot add covariation analysis as the metadata is quite incomplete and heterogeneous and likely to give misleading results - hence this was not performed.

See referee reports

Introduction

Increased depression rates have been frequently reported on patients with inflammatory bowel disease (IBD) (Graff *et al.*, 2009), which is a big concern from a clinical standpoint, since increased levels of stress and anxiety are major drivers of IBD relapse and severity (Mawdsley & Rampton, 2006). Both IBD and depression are heavily influenced by the gut microbiome structure, which controls anti-inflammatory processes and permeability in the gut, and communicates with the brain by a complex and close relationship with the Autonomic Nervous System that is known as the brain-gut axis (Foster & McVey Neufeld, 2013; Luna & Foster, 2015; Martin *et al.*, 2018).

Altered microbiomes can have big impacts on the health and development of both the gut and brain, and alterations in the ecology of this microbiome, a process known as dysbiosis, have been separately linked to both depression and IBD (Kaur *et al.*, 2011; Rogers *et al.*, 2016). While IBD has become one of the main focus on microbiome research for its clinical relevance and complex relationship with metabolic, immune and neurological processes (Huttenhower *et al.*, 2014), research on the effect of the microbiome in mental health are comparatively scarce, but have already shown promising results reducing anxiety and depression symptoms using probiotics (Bravo *et al.*, 2011; Pinto-Sanchez *et al.*, 2017). However, the relationship between different microbiome population structures and these conditions is still poorly understood.

The availability of the large amount of data derived from the recent explosion in metagenomics and metatranscriptomics provides unique opportunities for investigation. However, it is sometimes difficult to identify informative species. Recently, machine learning algorithms have been successfully applied because they allow the identification of patterns in situations where large, multi-dimensional and heterogeneous datasets are available.

Among the several machine learning approaches available, random forest is an algorithm used for classification and regression based on an ensemble that builds a population of decision tree classifiers, such that the result of a prediction

from a given set of features is the most frequent result from the different trees of the “forest” (Breiman, 2001). This is an efficient and generalist algorithm that has already been applied in several metagenomic investigations in human diseases, such as IBS (Saulnier *et al.*, 2011).

The aim of this work was to apply the random forest approach to identify the microbiome species that may be mostly involved in IBD and depression outcomes and that are responsible for the most relevant changes in the population structure between IBD, depression and patients comorbid for both conditions, and to provide insights on how the microbiome is involved in this comorbidity.

Methods

Database generation

The datasets used for the analyses were retrieved from the Inflammatory Bowel Disease Multi-Omics Database (IBDMDB) (Schirmer *et al.*, 2018), which is part of the Integrative Human Microbiome Project (NIH HMP Working Group *et al.*, 2009). The IBDMDB database contains a wide array of omics data (e.g., 16S and shotgun metagenomic, metatranscriptomic, proteomic and host genomes) of 132 individuals classified by IBD diagnostic in ulcerative colitis, Crohn's disease and controls. Participants provided bi-weekly stool samples at five hospitals in the United States. Metagenomic and metatranscriptomic data was processed as described in Schirmer *et al.*, 2018 (Abubucker *et al.*, 2012; Truong *et al.*, 2015).

Subject selection

From this dataset, the 70 unique participants who answered an additional self-reported depression and anxiety questionnaire during registration (the answers to which are listed in the HMP2 metadata, column EC to EL) were selected. As the questionnaire model was not specified, only individuals with raw scores over 6 on this test was considered as showing “signs of depression”. To calculate the raw scores, a severity scale was generated, with the following scores: 0, never; 1, rarely; 2, sometimes; 3, often; 4, always. The scores were then summed to give a final total. In the case of individuals undergoing multiple tests, the lower score was used. We selected a low threshold in order to be able to identify putative dysbiotic individuals that were not experiencing severe depression symptoms. All the others were classified as “no sign of depression”. The combination between the test and the IBD diagnosis divided the dataset in six groups: Crohn's disease with no detectable sign of depression (CD; n=15), Crohn's disease with signs of depression (CDD; n=20), ulcerative colitis with no sign of depression (UC; n=4), ulcerative colitis with signs of depression (UCD; n=11), signs of depression but no inflammation (nonIBDD; n=7) and the control group: no inflammation/no depression (nonIBD; n=13). As the experimental design of the IBDMDB consisted of a longitudinal study, each subject contributed several times to this study, and all the samples used for this analysis were sequenced by shotgun sequencing as described in Schirmer *et al.* The resulting datasets for metagenomic and metatranscriptomic consist of 1084 and 566 samples, respectively. The final tables after

pre-processing consist of 1486 columns, including Participant ID, data type, diagnostic, sex, mental score, and nested columns on the relative values of the different taxa.

Data analysis

For each of the six groups, abundance matrices of the metagenomic data, metatranscriptomic data, and the combination of metagenomics and metatranscriptomics were used for random forest classification. Each of the datasets was divided randomly into a training set (90% of the individuals) and a validation set (10% of the individuals). Random forest analysis was performed using the library *Scikit-learn* 0.19.1 (Pedregosa *et al.*, 2011) on the training sets to identify the most important species involved in discriminating the samples without losing predicting power. A 1000-fold cross-validation for the combined dataset, and 500-fold for metagenomic and metatranscriptomic data (see [Figure 1](#)), considering one model for each iteration was performed and only the most important species in the construction of this model was retained. Only models with a precision classification >80% were considered, and among the considered models, only species that appeared more in more than one were selected. Afterwards, the validation sets were run with the selected species only to measure the possible loss of predictive capability and computed the area under the receiver operating characteristic (auROC) curve for the prediction of the validation set classes as a performance metric.

Statistical analysis

In order to assess the significance of the differences between the abundances of the selected species, we performed a one-way ANOVA (*Scipy* 1.0.0, Jones *et al.*, 2001) with a Tukey's honest significant difference (HSD) post-hoc test. This test makes pair-wise comparisons between the different means to see which classes are different. For clarity, confidence intervals for Tukey's HSD test can be found in Supplementary Materials ([Supplementary Figure 1](#) and [Supplementary Figure 2](#)).

The functional activity of the selected species was retrieved from the HUMAnN metatranscriptomic analyses described above. Only the pathways in which the selected species are involved and those that were different between the groups from the ANOVA test were selected and the correlation between these species was calculated using Spearman's correlation coefficient. A significance level of 0.05 was applied for all statistical tests.

Results and discussion

Species selection and model validation

The random forest cross-validation selection of the most informative species showed a combined list of 24 species, as can be seen in [Figure 2](#). The validation models for DNA, RNA and the combined dataset shows micro-averaged auROC values of 0.96, 0.91 and 0.99, respectively ([Supplementary Figure 3–Supplementary Figure 5](#)). This small loss of information suggest a relevant role of the selected species in the interaction of both conditions, while the capability of the model to classify the validation data with with great accuracy shows that our model can generalize its results and it's not overfitting.

All species exhibited differences in at least one group in a one-way ANOVA ($\alpha=0.05$, [Supplementary Table 1](#)), and

no significant differences were found between DNA and RNA abundances for these species ([Supplementary Table 2](#)). This list of putative species pretends to be a trade-off between the all-relevant and minimal informative approaches. We chose this approach in order to get as broad of a list as possible while avoiding artifacts related to the longitudinal nature of the dataset.

In order to assess the effect of the small sample size of group UC, the same procedure was made grouping all samples with IBD together. As expected, we see some difference in the species selected. However, the species that showed stronger differences in the previous classification were also the stronger ones, with most of the species overlapping. The interesting exception is *Faecalibacterium prausnitzii* that was absent.

The non-dysbiotic microbiome

The analyses showed an increase in the number of species from the genus *Bacteroides* in dysbiotic groups compared with the control (nonIBD) ([Figure 3](#)), as has been reported in other dysbiotic samples (Bloom *et al.*, 2011), with the exception of *Bacteroides dorei*, which is more abundant in non-IBD than in any other group. Aside from *Bacteroides dorei*, nonIBD samples had a higher abundance of *Alistipes shahii* and *Ruminococcus bromii*, while a typical species associated with nonIBD, *Faecalibacterium prausnitzii*, was significantly decreased in nonIBDD and CD.

Crohn's disease abundance changes in depression

Both of the Crohn's disease-related groups (CD and CDD) showed higher abundances of *Bacteroides ovatus* and *Bacteroides uniformis*. However, CD samples exhibited higher abundances for several specific species, including *Bacteroides xylanisolvens*, *Parasutterella excrementihominis* and *Bacteroides fragilis*, compared with CDD, but decreased abundance of *Faecalibacterium prausnitzii*, which did not differ significantly in abundance between nonIBD and CDD groups.

Ulcerative colitis changes in depression

Ulcerative colitis samples had the most distinctive microbiome profile. Several species, including *Burkholderiales* bacterium 1_1_47, *Bacteroides eggerthii* and *Bacteroides finegoldii* were characteristic of this group, and absent in the others, except for *B. finegoldii*, which was also present in a lower abundance in nonIBD samples. Only UCD samples exhibited an increased abundance of *Bacteroides fragilis*, *Bacteroides vulgatus* and *Haemophilus pittmaniae*, this last species being almost exclusive to the UCD group.

Non-IBD changes in depression

The nonIBDD was the group with the highest number of changes in microbiome diversity when compared with its non-depressed counterpart ([Table 1](#)). However, most of those changes followed a similar pattern in other dysbiotic groups.

A notable change was observed in *Faecalibacterium prausnitzii*, which was present in almost the same abundances in nonIBD, UCD and CDD samples, and a high variability in UC while being significantly lower in CD and nonIBDD ([Supplementary Table 3](#) and [Supplementary Table 4](#)). This is

Figure 1. Workflow of the k-Fold Cross Validation approach for Random Forest. First the data gets split into train and validation sets (A). The train dataset will be iterated by the Cross Validation algorithm (B), while the validation set will be spared to test the model trained only with the reduced feature list (C).

Figure 2. Venn diagram for the species selected for each dataset.

Figure 3. DNA (**A**) and RNA (**B**) taxonomic abundances for the selected species. Abundances were quantified by the relative abundances of their sequences, and for each level they should sum to 1 (including unclassified sequences).

Table 1. Changes between Crohn's disease (CD), ulcerative colitis (UC) and control (nonIBD) in depressed compared with non-depressed subjects. Increases/decreases shown are statistically significant.

Species	CD	UC	nonIBD
<i>Alistipes shahii</i>	-	-	Increase
<i>Bacteroides ovatus</i>	-	-	Increase
<i>Subdoligranulum</i> sp.	-	Decrease	-
<i>Bacteroides xylanisolvens</i>	Decrease	-	Increase
<i>Parasutterella excrementihominis</i>	Decrease	-	-
<i>Burkholderiales</i> bacterium 1_1_47	-	Decrease	-
<i>Alistipes putredinis</i>	-	Decrease	Decrease
<i>Bacteroides stercoris</i>	-	-	Increase
<i>Faecalibacterium prausnitzii</i>	Increase	-	Decrease
<i>Bacteroides uniformis</i>	Decrease	-	Increase
<i>Bacteroides fragilis</i>	Decrease	Increase	-
<i>Lachnospiraceae</i> bacterium 7_1_58	Increase	-	-
<i>Bacteroides dorei</i>	-	-	Decrease
<i>Bacteroides vulgatus</i>	-	Increase	Increase
<i>Ruminococcus bromii</i>	-	-	Decrease
<i>Bacteroides finegoldii</i>	Decrease	Decrease	-
<i>Bacteroides eggerthii</i>	-	Decrease	Increase
<i>Parabacteroides goldsteinii</i>	-	-	Increase
<i>Haemophilus pittmaniae</i>	-	Increase	-

particularly interesting, since this species is considered to have anti-inflammatory activity. It seems counterintuitive to find a depleted population of one of the species most associated in the literature with a healthy microbiome compared to an IBD one in a group that doesn't show any inflammatory process. However, *Parabacteroides goldsteinii* was increased in non-IBDD and was depleted in all IBD groups in comparison with control samples. The *Parabacteroides* genus have been associated previously with anti-inflammatory activity (Neff *et al.*, 2016; Schirmer *et al.*, 2016), so the increase in abundance of this bacteria may explain why the nonIBDD microbiome is not associated with inflammation in the gut.

Other than *Parabacteroides goldsteinii*, nonIBDD samples did not contain other characteristic groups, and, more notably, none of the selected species was specific for depressed or non-depressed phenotypes.

Microbial functional activity

Regarding the functional activity of these species, seven pathways that were more abundant in dysbiotic groups than in nonIBD were identified (Supplementary Figure 1) and were correlated between each other and inversely correlated with

most of the others (Supplementary Figure 2 and Supplementary Table 5). Those pathways are folate transformations II, N10-formyl-tetrahydrofolate biosynthesis, *de novo* L-ornithine biosynthesis, superpathway of pyridoxal 5'-phosphate biosynthesis and salvage, phosphopantothenate biosynthesis I, preQ0 biosynthesis and queuosine biosynthesis. Folate (vitamin B9) and pyridoxal 5'-phosphate (vitamin B6) deficiencies have been linked both to depression (Coppen & Bolander-Gouaille, 2005; Hvas *et al.*, 2004; Mitchell *et al.*, 2014), as they are key for the synthesis of several neurotransmitters, and IBD (Pan *et al.*, 2017; Yakut *et al.*, 2010), although this association is not well understood and does not seem to be evidence of causation. Increased levels of L-ornithine derivatives have also been linked to depression (Zheng *et al.*, 2010). However, even if nonIBDD have the highest activity for almost all of these pathways, CD and UC were also significantly increased, while functional activity in CDD was generally lower and non-significant in some pathways. Moreover, UCD did not differ from nonIBD in any of them.

This difference in functional activity again highlights the lack of a concrete pattern of gut microbiome abundance between depressed groups.

Conclusions

The random forest approach was able to successfully identify informative changes in abundance at the species level, revealing specific patterns for the depressed and non-depressed groups without losing predictive power. We believe that this approach, and Machine Learning in general, can be really useful in a field of research where high dimensionality is always an issue.

This work provided, to our knowledge for the first time, an overview about the difference in the bacterial communities of patients with signs of depression and the combination with depression and inflammatory bowel disease. Our findings suggest a complex landscape of microbiome interactions, both at population structure and functional activity levels. However, the results showed that there are distinct taxonomic profiles within patients of IBD depending on their depression status, providing further input for future investigations.

Data availability

The datasets used for the analyses were retrieved from the [Inflammatory Bowel Disease Multi-Omics Database \(IBDMDB\)](#) (Schirmer *et al.*, 2018), a part of the Integrative Human Microbiome Project (NIH HMP Working Group *et al.*, 2009).

Grant information

Authors thank funding from The Danish Independent Research Council (Technology and Production) grant number DFF – 6111-00471. P.M.M. thanks funding from Technical University of Denmark, DTU Bioinformatics for his research assistantship to conduct this study.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Supplementary material

Supplementary Figure 1. Relative abundances of the pathways that showed significant differences between groups (alpha= 0.05).

[Click here to access the data.](#)

Supplementary Figure 2. Correlation between the different pathways contributed by the selected species. Color gradient shows positive (red) or negative (blue) correlation.

[Click here to access the data.](#)

Supplementary Figure 3. Receiver operating characteristic curves for the validation model with combined metagenomic and metatranscriptomic data.

[Click here to access the data.](#)

Supplementary Figure 4. Receiver operating characteristic curves for the validation model with metagenomic data.

[Click here to access the data.](#)

Supplementary Figure 5. Receiver operating characteristic curves for the validation model with metatranscriptomic data.

[Click here to access the data.](#)

Supplementary Table 1. ANOVA results for each of the selected species in metagenomic and metatranscriptomic data sets.

[Click here to access the data.](#)

Supplementary Table 2. A t-test was used to assess the difference between DNA and RNA abundances per species and a nested column per group.

[Click here to access the data.](#)

Supplementary Table 3. Tukey's honest significant difference test for the metagenomic data. Results are organized by species with two nested columns, confidence intervals at 0.95 and the decision. Each row represents a pair-wise comparison.

[Click here to access the data.](#)

Supplementary Table 4. Tukey's honest significant difference test for the metatranscriptomic data. Results are organized by species with two nested columns, confidence intervals at 0.95 and the decision. Each row represents a pair-wise comparison.

[Click here to access the data.](#)

Supplementary Table 5. Tukey's honest significant difference test for the pathways correlated pathways. Results are organized by species with two nested columns, confidence intervals at 0.95 and the decision. Each row represents a pair-wise comparison.

[Click here to access the data.](#)

References

- Abubucker S, Segata N, Goll J, *et al.*: **Metabolic reconstruction for metagenomic data and its application to the human microbiome.** *PLoS Comput Biol.* 2012; **8**(6): e1002358.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Bloom SM, Bijanki VN, Nava GM, *et al.*: **Commensal *Bacteroides* species induce colitis in host-genotype-specific fashion in a mouse model of inflammatory bowel disease.** *Cell Host Microbe.* 2011; **9**(5): 390–403.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Bravo JA, Forsythe P, Chew MV, *et al.*: **Ingestion of *Lactobacillus* strain regulates emotional behavior and central GABA receptor expression in a mouse via the vagus nerve.** *Proc Natl Acad Sci U S A.* 2011; **108**(38): 16050–5.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Breiman L: **Random Forests.** *Mach Learn.* 2001; **45**(1): 5–32.
[Publisher Full Text](#)
- Coppen A, Bolander-Gouaille C: **Treatment of depression: time to consider folic acid and vitamin B12.** *J Psychopharmacol.* 2005; **19**(1): 59–65.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Foster JA, McVey Neufeld KA: **Gut-brain axis: how the microbiome influences anxiety and depression.** *Trends Neurosci.* 2013; **36**(5): 305–312.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Graff LA, Walker JR, Bernstein CN: **Depression and anxiety in inflammatory bowel disease: a review of comorbidity and management.** *Inflamm Bowel Dis.* 2009; **15**(7): 1105–1118.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Huttenhower C, Kostic AD, Xavier RJ: **Inflammatory bowel disease as a model for translating the microbiome.** *Immunity.* 2014; **40**(6): 843–854.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Hvas AM, Juul S, Bech P, *et al.*: **Vitamin B₆ level is associated with symptoms of depression.** *Psychother Psychosom.* 2004; **73**(6): 340–343.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Jones E, Oliphant T, Peterson P: **{SciPy}: open source scientific tools for {Python}.** 2001.
[Reference Source](#)
- Kaur N, Chen CC, Luther J, *et al.*: **Intestinal dysbiosis in inflammatory bowel disease.** *Gut Microbes.* 2011; **2**(4): 211–216.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Luna RA, Foster JA: **Gut brain axis: diet microbiota interactions and implications for modulation of anxiety and depression.** *Curr Opin Biotechnol.* 2015; **32**: 35–41.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Martin CR, Osadchiy V, Kalani A, *et al.*: **The Brain-Gut-Microbiome Axis.** *Cell Mol Gastroenterol Hepatol.* 2018; **6**(2): 133–148.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Mawdsley JE, Rampton DS: **The role of psychological stress in inflammatory bowel disease.** *Neuroimmunomodulation.* 2006; **13**(5–6): 327–336.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Mitchell ES, Conus N, Kaput J: **B vitamin polymorphisms and behavior: evidence of associations with neurodevelopment, depression, schizophrenia, bipolar disorder and cognitive decline.** *Neurosci Biobehav Rev.* 2014; **47**: 307–320.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Neff CP, Rhodes ME, Arnolds KL, *et al.*: **Diverse Intestinal Bacteria Contain Putative Zwitterionic Capsular Polysaccharides with Anti-inflammatory Properties.** *Cell Host Microbe.* 2016; **20**(4): 535–547.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- NIH HMP Working Group, Peterson J, Garges S, *et al.*: **The NIH Human Microbiome Project.** *Genome Res.* 2009; **19**(12): 2317–2323.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Pan Y, Liu Y, Guo H, *et al.*: **Associations between Folate and Vitamin B12 Levels and Inflammatory Bowel Disease: A Meta-Analysis.** *Nutrients.* 2017; **9**(4): pii: E382.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Pedregosa F, Varoquaux G, Gramfort A, *et al.*: **Scikit-learn: Machine Learning in Python.** *J Mach Learn Res.* 2011; **12**: 2825–2830.
[Reference Source](#)
- Pinto-Sanchez MI, Hall GB, Ghajar K, *et al.*: **Probiotic *Bifidobacterium longum* NCC3001 Reduces Depression Scores and Alters Brain Activity: A Pilot Study in Patients With Irritable Bowel Syndrome.** *Gastroenterology.* 2017; **153**(2): 448–459.e8.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Rogers GB, Keating DJ, Young RL, *et al.*: **From gut dysbiosis to altered brain function and mental illness: mechanisms and pathways.** *Mol Psychiatry.* 2016; **21**(6): 738–748.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Saulnier DM, Riehle K, Mistretta TA, *et al.*: **Gastrointestinal microbiome signatures of pediatric patients with irritable bowel syndrome.** *Gastroenterology.* 2011; **141**(5): 1782–1791.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Schirmer M, Franzosa EA, Lloyd-Price J, *et al.*: **Dynamics of metatranscription in the inflammatory bowel disease gut microbiome.** *Nat Microbiol.* 2018; **3**(3): 337–346.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Schirmer M, Smeekens SP, Vlamakis H, *et al.*: **Linking the Human Gut Microbiome to Inflammatory Cytokine Production Capacity.** *Cell.* 2016; **167**(4): 1125–1136.e8.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Truong DT, Franzosa EA, Tickle TL, *et al.*: **MetaPhlAn2 for enhanced metagenomic taxonomic profiling.** *Nat Methods.* 2015; **12**(10): 902–903.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Yakut M, Ustün Y, Kabaçam G, *et al.*: **Serum vitamin B₁₂ and folate status in patients with inflammatory bowel diseases.** *Eur J Intern Med.* 2010; **21**(4): 320–323.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Zheng S, Yu M, Lu X, *et al.*: **Urinary metabonomic study on biochemical changes in chronic unpredictable mild stress model of depression.** *Clin Chim Acta.* 2010; **411**(3–4): 204–209.
[PubMed Abstract](#) | [Publisher Full Text](#)

Open Peer Review

Current Peer Review Status: ? ?

Version 2

Reviewer Report 27 June 2019

<https://doi.org/10.5256/f1000research.20778.r47383>

© 2019 Suhail Y. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Yasir Suhail

Department of Biomedical Engineering, Yale University, Storrs, CT, USA

I think the comments raised in the previous review have not been sufficiently addressed. Perhaps a point by point discussion and reply by the authors will be helpful. Specifically:

1. It is still not clear as to what is meant by 1000-fold CV on a data set with 70 samples of ~ 1400 features each. If I understand correctly based on the added text and discussion by Reviewer 2, since each sample had longitudinal data and ~ 1400 features, perhaps the longitudinal data was split up into multiple samples? If this is so, please specify clearly how this was done and what the new dimensions of the data set were, instead of 70 X 1486.
2. Supplementary Figures 1 and 2 still link to ROC curves rather than pathway data, as written in the text.
3. The [link](#) to HMP2 metadata in the article, under Methods->Subject Selection gives a 404 error.

Again, a point by point rebuttal of the comments from the first and second round of reviews, along with any appropriate modification to the article, will be helpful for readers and for reviewing the article.

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 26 June 2019

<https://doi.org/10.5256/f1000research.20778.r47384>

© 2019 Narayanan M. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Manikandan Narayanan

Department of Computer Science and Engineering (CSE) and Initiative for Biological Systems Engineering (IBSE), Robert Bosch Centre for Data Science and Artificial Intelligence (RBC-DSAI), Indian Institute of Technology (IIT) Madras, Chennai, Tamil Nadu, India

It is great that authors have attempted to address major concerns from both reviewers, with additional analysis, figure and clarifications to text.

A couple follow-up points regarding the additions in version 2 remain to be addressed (as explained below); furthermore, some issues that are minor yet important for increasing the readability/impact of the manuscript, which have been already mentioned in previous reviewer reports, remain to be addressed. I believe all of these changes can be incorporated into a new version without any additional analyses, and instead only with clarifications to text/figures. Readers would find a new version incorporating the suggested changes below much more valuable, and I look forward to reading this new version and associated point-by-point response to all reviewer reports posted so far.

Regarding the new additions in version 2 compared to version 1, I've the following comments:

1. It is good to know that the species with stronger differences continue to be the informative species when IBD samples are grouped together, but it would be valuable for readers to know the exact list of informative species with the grouped IBD samples. Hence, please provide a suppl. table of all species identified using this combined IBD samples, so that the readers can learn more about which species with stronger differences got replicated, find other interesting exceptions, etc.
2. The authors say that "This list of putative species pretends to be a trade-off between the all-relevant and minimal informative approaches....". Did the authors try out the all-relevant approach (i.e., classic t-test or ANOVA for each species)? If so, please provide the selected species from this all-relevant approach as a suppl. table. If not, please mention that other alternate approaches such as minimally informative or all-relevant approach to select species are also possible (and cite the Boruta paper¹ for more information on these alternate approaches), but not tried out.
3. Typos have crept into some of the newly added text. Please fix these:
 - "chose this approach >>ir<< order to get as broad of a list as possible"
 - "only species that appeared >>more in more than one<<", etc.

Regarding concerns already raised in version 1 reviewer reports (of both reviewer 1 and 2), I would like to give a few example issues that were not addressed:

1. The suppl figs 1-5 captions seems mixed up, as already reported by reviewer 1 in his report, and it doesn't appear to be fixed in version 2 (similarly the very small font size in this figure making it very hard to read has also not been fixed). While this is a minor issue that can be

easily fixed, leaving it unfixed can negatively impact the readability of the article.

2. The 1000-fold cross-validation still needs some explanation, again as raised by reviewer 1 in his response. Figure 1 helps a lot to understand the data splits, and with additional text on longitudinal sampling of the same individual, it is easier to understand that 70 individuals actually give rise to ~1400 data points. What is not clear are:
 - How are these ~1400 data points split into 1000 folds? Please clarify in text.
 - Please also clarify in text any issues/caveats associated with doing cross-validation on samples that are not independently distributed but are instead correlated due to several data points coming from the same individual.
3. There are several other, "simple-to-address" issues (i.e., issues that require no additional analyses, only clarifications to text/figures, to address), raised by reviewer 1 in his report. While a single paragraph summarizing all key changes to version 2 compared to version 1 is valuable, a point-by-point response to all reviewer reports submitted so far specifying which issues were addressed and which issues were beyond the scope of this work to address would make it easier for the reviewers to understand the reasoning of the authors in deciding which issues they decided to address when they prepared version 2.

A quantitative data-driven analysis of the brain-gut-microbiome axis is an important topic to understand, and this paper makes a significant contribution in this area, and it would be valuable to readers once the above comments are addressed.

References

1. Kursa M, Rudnicki W: Feature Selection with the Boruta Package. *Journal of Statistical Software*. 2010; **36** (11). [Publisher Full Text](#)

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Computational biology, bioinformatics, systems biology, gene networks, disease-disease interactions, multi-tissue genomics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Version 1

Reviewer Report 04 October 2018

<https://doi.org/10.5256/f1000research.16435.r37362>

© 2018 Narayanan M. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Manikandan Narayanan

¹ Department of Computer Science and Engineering (CSE) and Initiative for Biological Systems Engineering (IBSE), Robert Bosch Centre for Data Science and Artificial Intelligence (RBC-DSAI), Indian Institute of Technology (IIT) Madras, Chennai, Tamil Nadu, India

² Department of Computer Science and Engineering (CSE) and Initiative for Biological Systems Engineering (IBSE), Robert Bosch Centre for Data Science and Artificial Intelligence (RBC-DSAI), Indian Institute of Technology (IIT) Madras, Chennai, Tamil Nadu, India

Review process:

The open review model calls for a special reviewing strategy since previous reviews/comments are already available. I read the paper and formed my independent opinions before looking at the previous reviews. Then I wrote this report to express my opinions in the context of the previous reviewer's opinion.

Summary:

The authors attempt to unravel the complex interaction between IBD and depression that is potentially mediated by the gut microbiome. More specifically, they identify a subset of (24) microbial species that could discriminate between various single-disease and co-morbid cases of IBD and depression, and discuss the potential roles of these species and associated functional pathways in co-morbidity in the context of prior literature. Their methodology involves applying a combination of a machine learning approach (building a random forest to predict disease labels from species abundances and feature selection of informative species in the final model) and performing statistical tests (ANOVA-based test to identify which of the 24 selected/informative species differ between pairwise groups of interest such as the IBD with vs. without depression groups).

Strengths:

The authors use a systematic data-driven approach to address an important question of which microbial species and which functional pathways are involved in the "microbiome-brain-gut axis" (or specifically which species/pathways discriminate between different groups of individuals such as IBD disease with vs. without depression). The idea of using a random forest to select informative species with predictive power and then doing all subsequent analysis with the selected species is an interesting strategy to deal with the heterogeneity in the data and small sample size per strata/group (though I've some concerns with this approach/idea as mentioned below, which may be addressed with some additional analysis). The nice summary of key results in Table 1 (of species whose abundances have changed between IBD with vs. without depression) and the related discussion of identified species in the context of prior literature on single/comorbid cases of IBD/depression would be of immediate interest to researchers in this field.

Concerns:

I agree with the previous reviewer's concerns/feedback on improving presentation (such as by adding key details on data dimensions and cross-validation folds to the text) and improving interpretation (such as by comparing the 24 selected species to whatever species would be detected from applying a classic t-test or ANOVA test to all species rather than just the 24 species selected in the random forest analysis).

I now provide my feedback/comments not already covered by the previous reviewer below.

1. Statistical concerns due to small sample size and potentially confounding covariates:

Though the overall population is a decent sample size of 70 individuals, the per-strata or per-group sample sizes are low to moderate, with some groups like UC having only 4 patients!! The authors are aware of this issue and use a random forest feature selection to tackle the heterogeneity in sparse human population data. But I am not fully convinced that a machine learning model can recover from insufficient sample sizes as low as $n=4$. One way to address this issue could be to merge together UC with CD and then label them as a single IBD group and then study this group with/without signs of depression. If the conclusions are similar before/after this merging of UC with CD, then the authors may mention that this additional test yielded similar results and keep the current results in the paper as is.

Another issue with heterogeneous and sparse human data is that covariates such as age, gender, BMI, genotype, etc. are more likely to confound the association of gut microbiota with disease status. Are these covariates available from the original cohort (I would assume so since the authors say that even host genomes are available)? Importantly, if available, are these covariates matched between the different groups being compared here? If not matched, is the data adjusted for these covariates and how do the results change before/after this adjustment? Providing such information would be critical to readers to properly interpret the detected microbiota associations with IBD/depression.

2. The all-relevant vs. minimal-informative set of species:

Please provide clarification on the text on whether the 24 species is a minimal, non-redundant set of species that has predictive power to classify the disease labels, or whether it includes all the relevant species that is associated with disease labels (or whether it is somewhere in-between in this spectrum). In other words, is any other species other than these 24 species associated with disease status? While a minimal set is sufficient to build a predictive model and simplifies further interpretation, the all-relevant set would be useful to understand the comprehensive role of all species and the overall mechanisms involved, as explained nicely in the Introduction of this paper¹ on the Boruta feature selection package.

Based on the details provided on random forest based feature selection, the reported results may be closer to the all-relevant than a minimal-informative set, but the requirement of a species to be present in at least 2 models to be selected as an informative species is somewhat ad hoc (i.e., why not 3 or 4 models as a cutoff) and makes it unclear on whether all relevant species are selected. A more systematic way would be to assess the statistical significance of each species' association using a "wrapper method" around the random forest, such as the shuffling-based Boruta feature selection package (which is also used in the Saulnier et al. 2011 paper that the authors have cited). An alternative could be one of the methods in the paper "Statistical interpretation of machine learning-based feature importance scores for biomarker discovery"².

3. The context from prior literature that the authors already provide could be strengthened even further by connecting them to the brain-gut-microbiome axis reviewed in some recent articles^{3,4}.

References

1. Kursa MB, Rudnicki WR: Feature selection with the Boruta package. *J Stat Softw.* 2010; **36** (11).
2. Huynh-Thu VA, Saeys Y, Wehenkel L, Geurts P: Statistical interpretation of machine learning-based feature importance scores for biomarker discovery. *Bioinformatics.* 2012; **28** (13): 1766-74 [PubMed Abstract](#) | [Publisher Full Text](#)
3. Martin CR, Osadchiy V, Kalani A, Mayer EA: The Brain-Gut-Microbiome Axis. *Cell Mol Gastroenterol Hepatol.* 2018; **6** (2): 133-148 [PubMed Abstract](#) | [Publisher Full Text](#)
4. Kelly JR, Clarke G, Cryan JF, Dinan TG: Brain-gut-microbiota axis: challenges for translation in psychiatry. *Ann Epidemiol.* **26** (5): 366-72 [PubMed Abstract](#) | [Publisher Full Text](#)

Is the work clearly and accurately presented and does it cite the current literature?

Partly

Is the study design appropriate and is the work technically sound?

Partly

Are sufficient details of methods and analysis provided to allow replication by others?

Partly

If applicable, is the statistical analysis and its interpretation appropriate?

Partly

Are all the source data underlying the results available to ensure full reproducibility?

Partly

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Computational biology, bioinformatics, systems biology, gene networks, disease-disease interactions, multi-tissue genomics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 25 June 2018

<https://doi.org/10.5256/f1000research.16435.r35101>

© 2018 Suhail Y. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Yasir Suhail

¹ Department of Biomedical Engineering, Yale University, Storrs, CT, USA

² Department of Biomedical Engineering, Yale University, Storrs, CT, USA

Overview

The paper's central thesis of the correlation, or putative causative mechanism of the gut microbiome on IBD and depression, is important. The application of machine learning techniques may be suitable because of the high dimensional structure of the data and lack of knowledge of a simple causal relationship between the microbiome and IBD/depression.

I have a number of concerns and suggestions about the article in its present form. Most of these concern the presentation of the data, methods, and results. Addressing these concerns will help bring the article into a shape where readers can better evaluate the results and contribution of this study.

Presentation of the methods and results

1. The article should explicitly specify the dimensions and form of the data/features that were fed to the random forest algorithm. Were they the estimates of each species abundance in each individual from the respective RNA and DNA sequences? Was it derived from ribosomal sequences only? How many species were included? What was the total dimension of this data set? Does it correspond to a specific table in the IBDMDB resource?
2. Since the intended audience may be researchers generally interested in IBD or metagenomics who are not experts in random forests, algorithmic details should be provided in terminology that is common to the broader field. N-fold cross-validation generally refers to using 1 out of N data samples as a hold out test sample. However, the article mentions 1000-fold and 500-fold cross-validation for a data-set with 70 individuals. Does this correspond to individual re-sampled bagging or feature bagging in the random forest? A more explicit explanation will make this comprehensible to more readers.
3. The supplementary figures and tables may have been mixed/corrupted. S. Figures 1 and 2 are described as corresponding to pathway analysis but are actually ROC curves. SF4 and 5 are supposed to be ROC curves but probably show pathway results. The resolution is too low to make out the text. The method of calculating the pathway abundances should also be described somewhere. Is it the total number of reads corresponding to genes in the pathway, does it depend on the species abundances or any other parameters?
4. Some of the ambiguity in the analysis may be removed by providing the code for any pre-processing, random forest analysis, feature selection, and pathway abundance analysis etc.

Interpretation and Conclusion

A machine learning algorithm can build an accurate prediction system, or generate hypotheses about the mechanisms at play or provide some other insight into the process. Here, I see two possible results of the ML analysis:

1. The prediction accuracy can be a measure of the amount of information contained in the microbiome about the diseases. Alternatively, how predictive is the gut microbiome, and does this imply evidence for the causative effect of the microbiome on the disease? These

would be comparatively harder claims to make, and would probably require a few more calculations.

2. The random forests are used to arrive at the most important features (bacterial species) affecting bowel disease and depression. I think this is the main claim/result of the analysis. In this case, how much more does information does ML give us compared to simply finding the species whose abundance is most different between the disease and non-disease states (in terms of fold-change or p-value). For the case of the multi-class problem, ANOVA can provide p-values for the non-random abundances in the different classes of patients. The article describes the results of such t-tests and ANOVA results. A sufficient and logical argument for the ML approach supported by any relevant calculations will strengthen the case for this analysis.

Overall, I feel the discussion of the possible role of some of the species and metabolic pathways etiology of the disease is the most interesting for biologists and clinicians. The article is important in this regard and further development of this discussion can only add to its strength.

Is the work clearly and accurately presented and does it cite the current literature?

No

Is the study design appropriate and is the work technically sound?

Partly

Are sufficient details of methods and analysis provided to allow replication by others?

No

If applicable, is the statistical analysis and its interpretation appropriate?

Partly

Are all the source data underlying the results available to ensure full reproducibility?

Partly

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 17 Sep 2018

Pedro Morell Miranda

The precision metric was used for each Cross Validation model, so each precision metric was calculated using the test split for that iteration from the original train dataset. The

validation dataset was only used to measure the final metrics of the model trained only with the features selected by the Cross Validation method.

Regarding the "only species that appeared more than once were selected" statement, it refers to the fact that, from the list of selected features, we saw a few cases of species present only in all the samples of only one individual, which in some cases can end being considered as quite characteristic by the model as each patient contributed with several samples. Setting this condition allowed us to avoid those "one hit wonders" that were not descriptive of the group and appeared only because of the longitudinal experimental design.

Competing Interests: No competing interests were disclosed.

Comments on this article

Version 2

Reviewer Response 01 May 2019

Yasir Suhail

I see the new figure added in Version 2, but I am still not sure how 1000-fold cross validation was performed on a dataset of 70 individuals. Even if only 1 of the samples is used as a test set, this implies a 70-fold cross validation. Or does 1000 fold cross validation mean something else in this study? If so, it should be specified clearly what is meant by this cross-validation here.

Competing Interests: No competing interests were disclosed.

Version 1

Reviewer Response 25 Jun 2018

Yasir Suhail

Expanding on Prof. Waldron's comment, the article states "Only models with a precision classification >80% were considered". If this precision is evaluated on the test/validation set then the validation set has already been used for training, biasing the estimate of accuracy.

I think part of the confusion might be phrases such as "running validation sets", or "validation models". If these refer to using the random forest models trained only on the training set to predict the disease labels on the test set, everything is fine. But people might understand a "validation model" to mean multiple things.

In the same vein, the article states "and among the considered models, only species that appeared more than once were selected". Does this mean that decision trees were pruned to only keep those that had these selected species as their initial steps? Or was a random forest re-trained using these selected species? In this case, either it was trained using the same training set used initially, or it was trained on the validation set.

Competing Interests: No competing interests were disclosed.

Author Response 14 Jun 2018

Pedro Morell Miranda

Dear Prof. Waldron,

The "validation set" was only used for prediction. The validation model was still trained with the "training set". We did this in order to see how our model fits new data and to avoid the scenario described in Cawley & Talbot, 2010, in section 5.3, where we are testing on samples that the model has already "seen". As you correctly point out, training on a different dataset would leave us with no way to estimate the predictive power of our model.

Competing Interests: Author

Reader Comment 12 Jun 2018

Levi Waldron

Please note that by training a random forest model in your 10% "validation set", this becomes a second training set, leaving you no way to estimate the predictive accuracy of the model. See for example section "5. Bias in Performance Estimation" of Cawley GC, Talbot NLC: **On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation**. *J. Mach. Learn. Res.* 2010, **11**:2079–2107 or the discussion of "resubstitution" error in Molinaro AM, Simon R, Pfeiffer RM: **Prediction error estimation: a comparison of resampling methods**. *Bioinformatics* 2005, **21**:3301–3307.

Competing Interests: No competing interests were disclosed.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research