RESEARCH NOTE

Investigation of gut microbiome association with inflammatory bowel disease and depression: a machine learning approach [version 1; referees: awaiting peer review]

Pedro Morell Miranda, Francesca Bertolini, Haja N. Kadarmideen

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

v1 **First published:** 05 Jun 2018, **7**:702 (doi: 10.12688/f1000research.15091.1)
Latest published: 05 Jun 2018, **7**:702 (doi: 10.12688/f1000research.15091.1)

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 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.

Open Peer Review

Referee Status: *AWAITING PEER REVIEW*

Discuss this article

Comments (0)

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**: Supervision, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

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 1; referees: awaiting peer review]** *F1000Research* 2018, **7**:702 (doi: 10.12688/f1000research.15091.1)

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.

First published: 05 Jun 2018, **7**:702 (doi: 10.12688/f1000research.15091.1)

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 Autonomous Nervous System that is known as the brain-gut axis (Foster & McVey Neufeld, 2013; Luna & Foster, 2015).

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). However, little is known about the role of the microbiome in the two diseases.

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).

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 were 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, 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 than once 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 1. 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 metrics highlight the performance of the model that, even with a reduced subset of species, has not lost predictive power.

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).

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 2), as has been reported in other dysbiotic samples (Bloom *et al*., 2011), with the exception of *Bacteroides dorei*, which is more abundant in nonIBD 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 prausnitzi*i, 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.

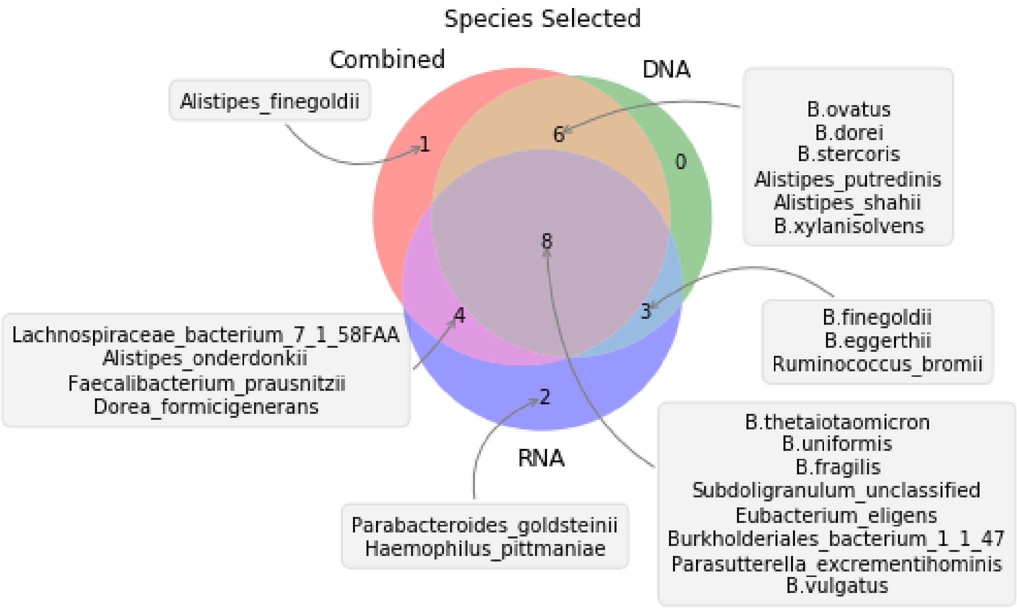

Figure 1. Venn diagram for the species selected for each dataset.

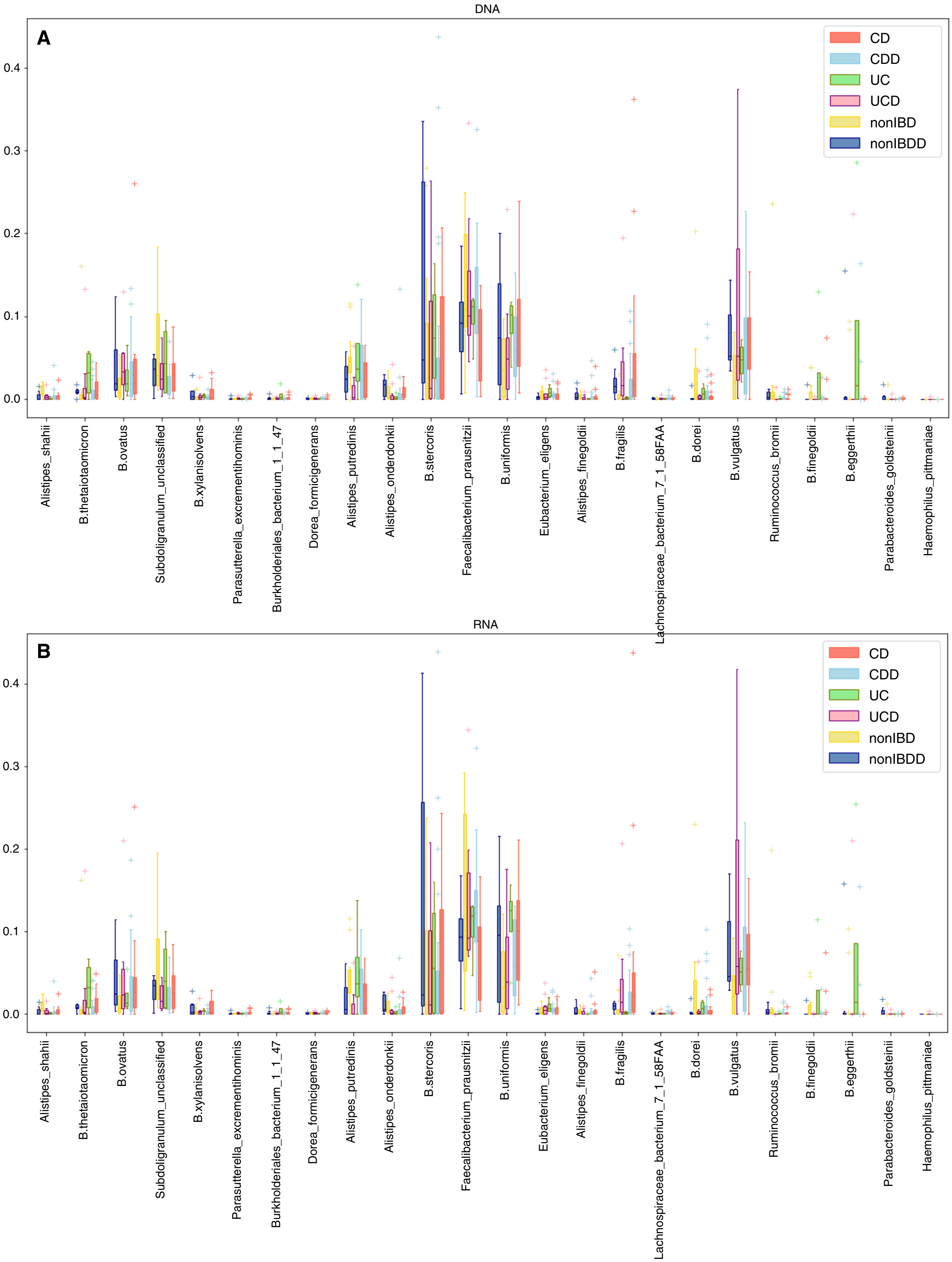

Figure 2. DNA (**A**) and RNA (**B**) taxonomic abundances for the selected species. Abundances were quantified by the relative abundances of their sequences, and for each level they should sum to 1 (including unclassified sequences).

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
Alistipes shahii	-	-	Increase
Bacteroides ovatus	-	-	Increase
Subdunigranulum sp.	-	Decrease	-
Bacteroides xylanisolvens	Decrease	-	Increase
Parasutterella excrementihominis	Decrease	-	-
Burkholderiales bacterium 1_1_47	-	Decrease	-
Alistipes putredinis	-	Decrease	Decrease
Bacteroides stercoris	-	-	Increase
Faecalibacterium prausnitzii	Increase	-	Decrease
Bacteroides uniformis	Decrease	-	Increase
Bacteroides fragilis	Decrease	Increase	-
Lachnospiraceae bacterium 7_1_58	Increase	-	-
Bacteroides dorei	-	-	Decrease
Bacteroides vulugatus	-	Increase	Increase
Ruminoccocus bromii	-	-	Decrease
Bacteroides finegoldii	Decrease	Decrease	-
Bacteroides eggerthii	-	Decrease	Increase
Parabacteroides goldsteinii	-	-	Increase
Haemophilus pittmaniae	-	Increase	-

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 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 nonIBDD and was depleted in all IBD groups in comparison with control samples. The *Parabacteroides* genre 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 pyroxidal 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. 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).

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.

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

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

Hvas AM, Juul S, Bech P, *et al.*: **Vitamin B_6 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

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

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 B_{12} 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