

Article

Gut Microbial Shifts Indicate Melanoma Presence and Bacterial Interactions in a Murine Model

Marco Rossi ^{1,†} , Salvatore M. Aspromonte ^{2,3,†} , Frederick J. Kohlhapp ³, Jenna H. Newman ³, Alex Lemenze ⁴, Russell J. Pepe ², Samuel M. DeFina ³ , Nora L. Herzog ³ , Robert Donnelly ⁴, Timothy M. Kuzel ¹ , Jochen Reiser ¹, Jose A. Guevara-Patino ^{5,*} and Andrew Zloza ^{1,*}

¹ Rush University Medical Center, Chicago, IL 60612, USA; marco_rossi@rush.edu (M.R.); timothy_kuzel@rush.edu (T.M.K.); jochen_reiser@rush.edu (J.R.)

² Rutgers Robert Wood Johnson Medical School, Rutgers, The State University of New Jersey, New Brunswick, NJ 08901, USA; saa9216@nyp.org (S.M.A.); rjpepe19@gmail.com (R.J.P.)

³ Rutgers Cancer Institute of New Jersey, Rutgers, The State University of New Jersey, New Brunswick, NJ 08901, USA; kohlhapp@gmail.com (F.J.K.); jenna.newman@mssm.edu (J.H.N.); sam.defina@yale.edu (S.M.D.); herzog.nora@gmail.com (N.L.H.)

⁴ Rutgers New Jersey Medical School, Rutgers, The State University of New Jersey, Newark, NJ 07103, USA; alemenze@gmail.com (A.L.); donnelly@rutgers.edu (R.D.)

⁵ Moffitt Cancer Center, Tampa, FL 33612, USA

* Correspondence: josealejandro.guevara@moffitt.org (J.A.G.-P.); andrew_zloza@rush.edu (A.Z.)

† These authors contributed equally to this work.



Citation: Rossi, M.; Aspromonte, S.M.; Kohlhapp, F.J.; Newman, J.H.; Lemenze, A.; Pepe, R.J.; DeFina, S.M.; Herzog, N.L.; Donnelly, R.; Kuzel, T.M.; et al. Gut Microbial Shifts Indicate Melanoma Presence and Bacterial Interactions in a Murine Model. *Diagnostics* **2022**, *12*, 958. <https://doi.org/10.3390/diagnostics12040958>

Academic Editor: Sung Chul Lim

Received: 4 February 2022

Accepted: 4 March 2022

Published: 12 April 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Through a multitude of studies, the gut microbiota has been recognized as a significant influencer of both homeostasis and pathophysiology. Certain microbial taxa can even affect treatments such as cancer immunotherapies, including the immune checkpoint blockade. These taxa can impact such processes both individually as well as collectively through mechanisms from quorum sensing to metabolite production. Due to this overarching presence of the gut microbiota in many physiological processes distal to the GI tract, we hypothesized that mice bearing tumors at extraintestinal sites would display a distinct intestinal microbial signature from non-tumor-bearing mice, and that such a signature would involve taxa that collectively shift with tumor presence. Microbial OTUs were determined from 16S rRNA genes isolated from the fecal samples of C57BL/6 mice challenged with either B16-F10 melanoma cells or PBS control and analyzed using QIIME. Relative proportions of bacteria were determined for each mouse and, using machine-learning approaches, significantly altered taxa and co-occurrence patterns between tumor- and non-tumor-bearing mice were found. Mice with a tumor had elevated proportions of *Ruminococcaceae*, *Peptococcaceae.g_rc4.4*, and *Christensenellaceae*, as well as significant information gains and ReliefF weights for *Bacteroidales.f__S24.7*, *Ruminococcaceae*, *Clostridiales*, and *Erysipelotrichaceae*. *Bacteroidales.f__S24.7*, *Ruminococcaceae*, and *Clostridiales* were also implicated through shifting co-occurrences and PCA values. Using these seven taxa as a melanoma signature, a neural network reached an 80% tumor detection accuracy in a 10-fold stratified random sampling validation. These results indicated gut microbial proportions as a biosensor for tumor detection, and that shifting co-occurrences could be used to reveal relevant taxa.

Keywords: gut microbiota; machine learning; statistical algorithms; co-occurrence patterns; melanoma

1. Introduction

The gastrointestinal microbiota contains a diverse and dense collection of symbiotic organisms that contribute to intestinal homeostasis. Nutrient digestion, synthesis of vitamins, protection against pathologic organisms, and production of neurotransmitters are just a few of the biological functions that these organisms provide [1–3]. The host's immune system plays an essential role in controlling microbial growth and development in the microbiome to ensure that a mutual relationship is maintained between the host and organism.

At the same time, the microbiota plays a role in adapting the host's immune system to various stressors [4]. In fact, evidence is accumulating that the intestinal microflora can respond to changes in host health status by sensing soluble host elements and local micro-environmental cues [5]. For this reason, the gastrointestinal microbiota is affected by the pathological immune responses derived from diseases such as diabetes mellitus, cancer, obesity, and inflammatory diseases, which impacts the body's immune response against disease [2,6,7].

It is increasingly being recognized that the gut microbiome composition differs significantly between healthy individuals and those with various pathological conditions. Dongmei et al. found that healthy individuals have a more diverse gut flora than those with colorectal cancer. In addition, certain bacterial populations were more likely to co-occur in patients with colorectal cancer than in healthy individuals [3]. While alterations in microbiome composition can be seen in pathologic conditions such as cancer, it is unclear whether these changes are a cause or a consequence of the disease [6]. Multiple studies that analyzed the composition of the gut microbiota in colorectal cancer patients suggested the presence of both "driver bacteria", or those that promote cancer growth, and "passenger bacteria", or those that solely flourish in the proinflammatory environment, but do not impact tumor progression. Geng et al. found that in their colorectal cancer patients, members of the *Enterobacteriaceae* family promoted cancer growth, whereas members of the *Streptococcaceae* family merely flourished in a proinflammatory environment [7].

The presence of these microbial mechanisms in which bacterial taxa have a certain level of dependency have wide implications for their use in modeling respective pathological conditions. Typically, connectivity and dependency between variables such as bacterial taxa in the context of predictive modeling has typically been a hindrance to model performance [8–10]. It is widely understood with many kinds of algorithms that, in various circumstances, variables with some manner of co-occurrence provide a certain level of redundant information, and therefore reduce the variability explained in models [8]. This presence of redundant information decreases the model's fit to the training dataset, as well as its prediction accuracy in the testing dataset [10–12].

Despite these limitations, co-occurrences in the context of pathological prediction with microbial taxa may still hold significance in the application of diagnostic signatures [8,13]. When co-occurrences shift between conditions, so does the direction of variability represented by relevant taxa in planes of higher dimensionality [9,10,14]. These shifts are reflected in principal component analysis, in which each principal component represents a different proportion of the total variability present [8,13]. They are also represented in ReliefF and information gain values, in which microbial taxa with these differences in variability have increased reliability as predictors [11,15]. Therefore, the identification of these shifts in co-occurrences in pathological conditions such as cancer is optimal for the implementation of gut microbial diagnostic signatures.

The implementation of machine-learning algorithms for the prediction of the presence of various cancers using the gut microbiome has been widely studied [16–18]. However, to date, relatively little work has been done regarding the use of the gut microbiome to predict the presence of melanoma. In addition, one of the challenges of predicting the presence of a specific disease with the gut microbiota is the variability in relative proportions of specific gut bacteria that can exist between patients and populations [12]. Through our analyses, we have indicated shifts in microbial co-occurrences as a potential method in accounting for such variability. Therefore, we hypothesized that models based on gut microbial proportion profiles of taxa involved in co-occurrence shifts could form a distinct diagnostic signature that effectively differentiated mice bearing mouse melanoma tumors from non-tumor-bearing mice. This implies that the intestinal microflora may function as a biosensor for the presence of cancer, and that its manipulation may alter cancer prognoses.

2. Results

2.1. Shifts in Microbial Taxon Proportions of Melanoma-Bearing Mice

Mice bearing melanoma tumors displayed significant shifts in gut microbial proportions compared to non-tumor-bearing mice, which: (1) implicated consistency in changes in gut microbiota data with tumors in the skin, distal to the gut; and (2) implied that such changes could be used by an algorithm to detect the presence of cancer. We compared the microbial composition of fecal samples of melanoma-bearing and tumor-free mice by terminal restriction fragment length polymorphism (T-RFLP) analysis [14,16]. This technique is commonly used to study complex microbial communities based on 16S rRNA gene variation, and has been applied in the study of microbial communities in soil and sludge systems [19]. T-RFLP analysis was carried out in a blinded fashion as previously described [4]. It was readily seen for the two mouse experiments (Figure 1) that the co-occurrences of relative taxon proportions shifted in the presence of B16 melanoma. In addition, *Peptococcaceae.g_rc4.4* was significantly increased (Wilcoxon $p < 0.05$) in both groups of mice (Figure 1). These data demonstrated that the intestinal flora developed detectable changes that discriminated a tumor-bearing from a tumor-free host. In order to more fully determine the extent to which these results distinguished between hosts that had a tumor and those that did not, the two mouse groups were combined and further analyzed as a single dataset ($n = 56$).

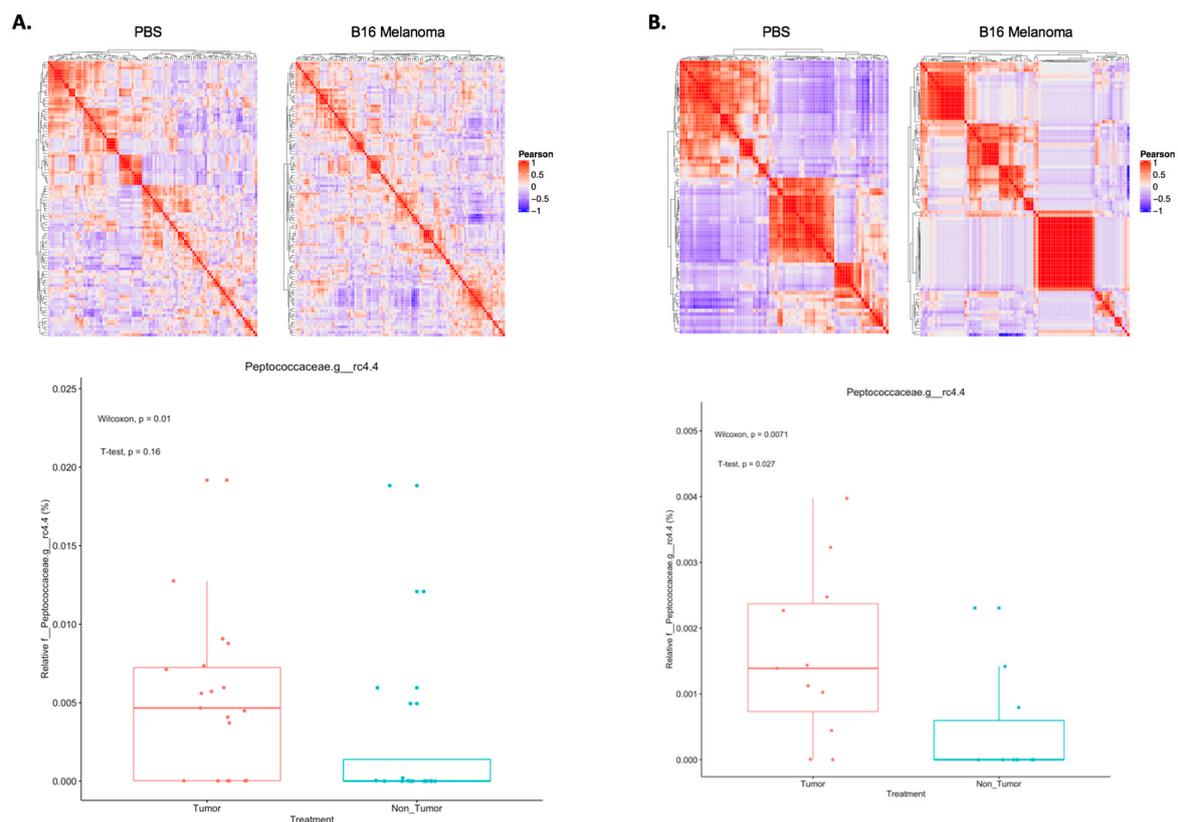


Figure 1. Shifted co-occurrences of microbial taxa and increased *Peptococcaceae.g_rc4.4* characterize tumor presence. (A) C57BL/6 (B6) male mice were injected with either 10^5 B16 melanoma cells ($n = 19$) or PBS ($n = 16$). After 10 days, fecal samples were collected and 16S rRNA genes were analyzed using terminal restriction fragment length polymorphism (T-RFLP) analysis. From individual taxon proportion and co-occurrence patterns, it could be seen that such patterns shifted with melanoma presence, and *Peptococcaceae.g_rc4.4* levels increased. (B) B6 male mice were injected with either 10^5 B16 melanoma cells ($n = 11$) or PBS ($n = 10$). After 16 days, fecal samples were collected and 16S rRNA genes were analyzed using terminal restriction fragment length polymorphism (T-RFLP) analysis. The results of these data directly corresponded with the mice in (A).

2.2. Co-Occurrence between *Bacteroidales.f__S24.7*, *Clostridiales*, and *Ruminococcaceae* Proportions in Mouse Melanoma

Seeking to identify the specific bacterial co-occurrences that were altered in the presence of a tumor, we first used Cytoscape to map them in the B16-melanoma- and PBS-treated mice. From these diagrams (Figure 2A,B), it was found that the co-occurrences of *Bacteroidales.f__S24.7* greatly differed between the two treatments. When looking further into this taxon, it was found that its co-occurrences with *Clostridiales* and *Ruminococcaceae* had changed the most between tumor and nontumor/PBS (Figure 2C,D), with Pearson correlation values of approximately -0.9 and -0.8 for tumor, as well as -0.15 and -0.13 for nontumor, respectively. Interestingly, however, when looking at the individual relative amounts of these taxa, the only one that was significantly different between tumor and nontumor was *Ruminococcaceae* (Wilcoxon $p < 0.05$, T -test $p < 0.05$; Figure 2E). Thus, we concluded that the potential for these taxa to predict tumor presence relied heavily on the extent to which their co-occurrences shifted in that condition, rather than changes in their individual relative amounts.

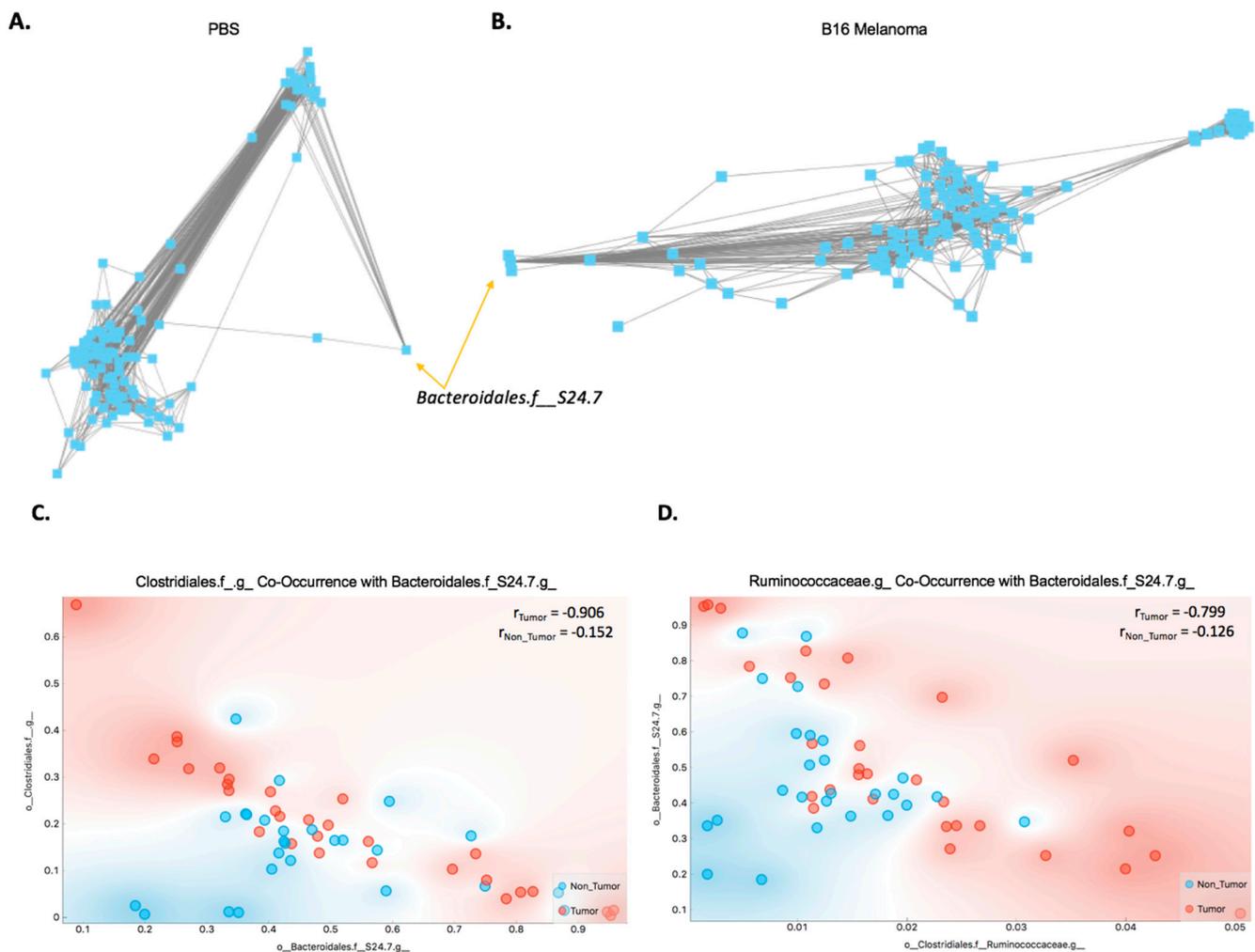


Figure 2. Cont.

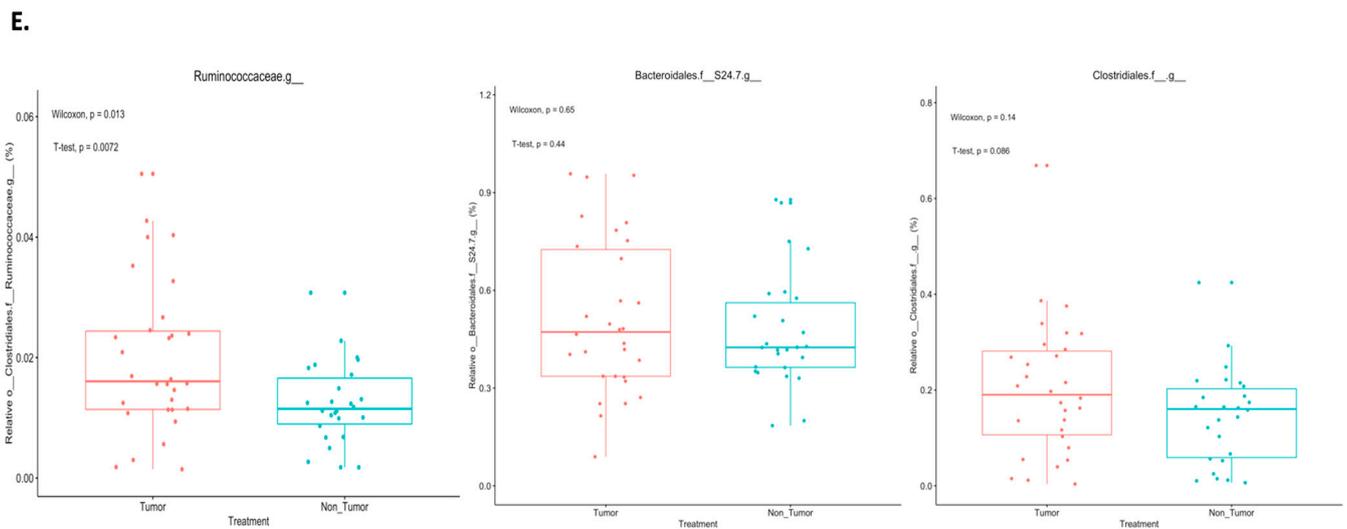


Figure 2. Co-occurrence changes between *Bacteroidales.f__S24.7*, *Clostridiales*, and *Ruminococcaceae* occur with tumor presence. (A,B) Pearson correlation matrices were determined for microbiotas from tumor and nontumor mice and displayed using Cytoscape. From these visualizations, *Bacteroidales.f__S24.7* co-occurrences greatly changed with tumor presence. (C,D) Using the R programming language, it was found that the most dramatic shifts of *Bacteroidales.f__S24.7* were in conjunction with *Clostridiales* and *Ruminococcaceae*. (E) When comparing each taxon individually between tumor and nontumor, only *Ruminococcaceae* was significantly different.

2.3. Differences in Principal Components between Tumor and Nontumor

Considering our results for both individual microbial taxa and co-occurrence shifts, we wanted to assess the relevance of each taxon in the context of predictive modeling. Thus, we calculated the information gains and ReliefF weights for each taxon (Figure 3A,B). In the scoring for information gains, *Ruminococcaceae*, *Peptococcaceae.g_rc4.4*, and *Christensenellaceae* consistently scored higher than the majority of taxa (Figure 3A). For the ReliefF algorithm, *Bacteroidales.f__S24.7* had a fairly high weight, along with *Peptococcaceae.g_rc4.4* and *Christensenellaceae* (Figure 3A). Further, *Christensenellaceae* was found to be significantly different between tumor and nontumor (Wilcoxon $p < 0.05$, Figure 3A,B). Considering that *Bacteroidales.f__S24.7* shifted its co-occurrences and its ReliefF weight indicated variable importance, we performed a principal component analysis (PCA) using this taxon (Figure 3C,D). Two PCAs were performed, one with *Clostridiales* and the other with *Ruminococcaceae* (Figure 3C,D). After performing the PCAs, we compared the resulting principal component coordinates between tumor and nontumor mice. From this comparison, we found that, although the first principal components did not differ between the two groups (Figure 3C), the second ones did (Wilcoxon $p < 0.05$, T-test $p < 0.05$; Figure 3D). These results indicated that the coordinates of these second principal components could be implemented in predictive modeling.

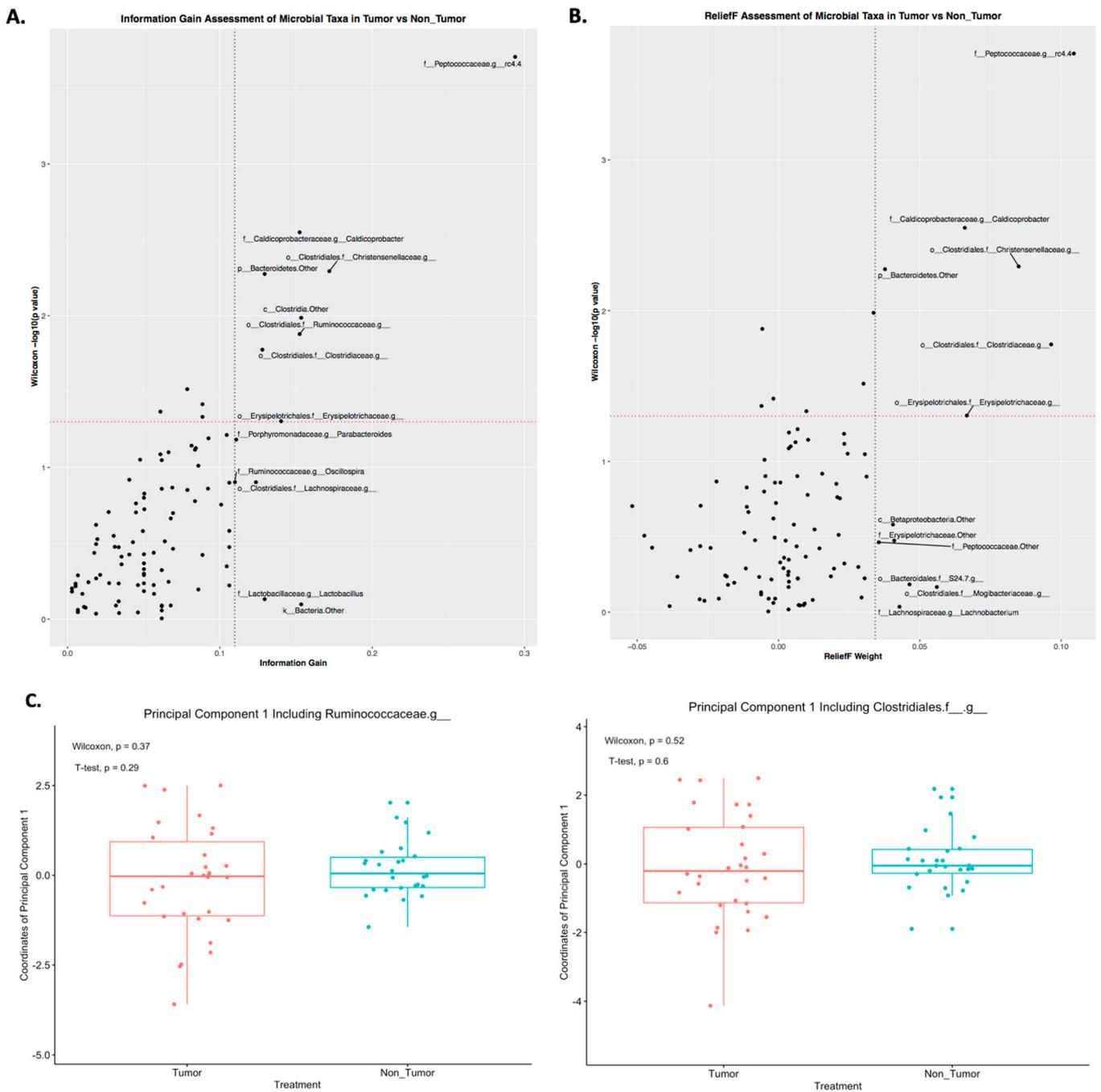


Figure 3. Cont.

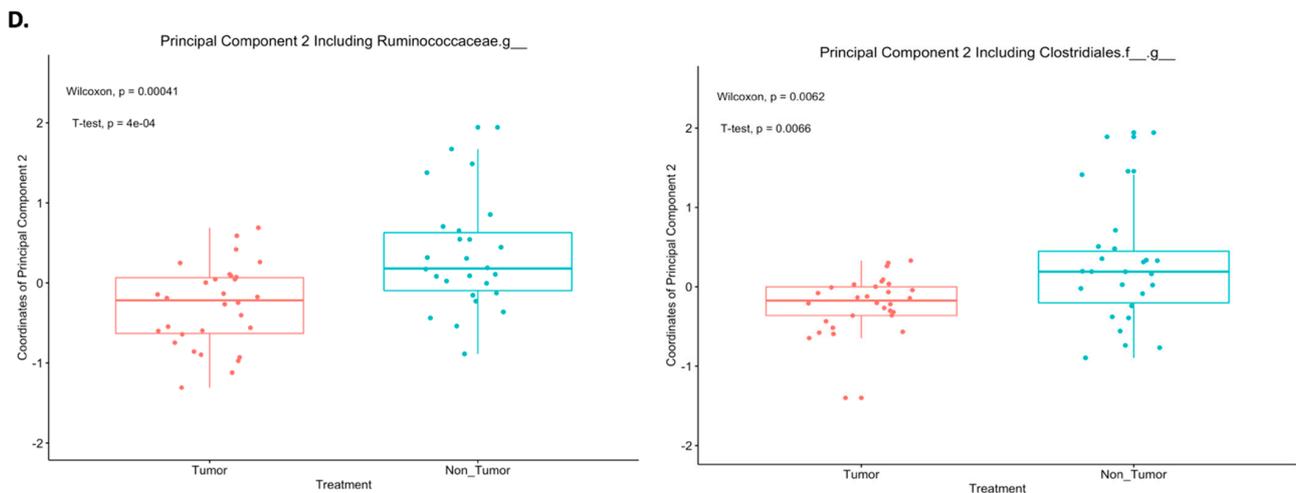
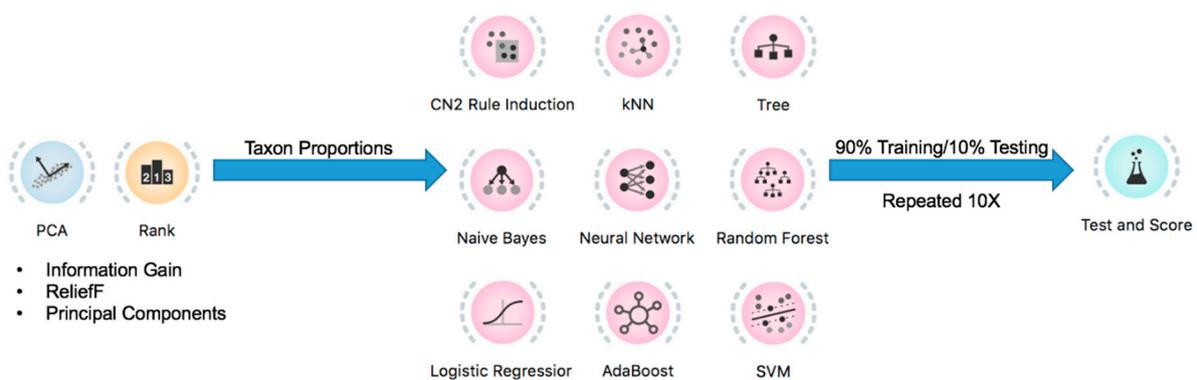


Figure 3. Significant predictors of tumor presence include the second principal components involving *Bacteroidales.f__S24.7*, *Clostridiales*, and *Ruminococcaceae*. (A,B) Using the CORElearn package in the R programming language, the information gains and ReliefF weights were calculated for each taxon. (A) *Ruminococcaceae*, *Peptococcaceae.g_rc4.4*, and *Christensenellaceae* were found significantly altered with tumor presence and having high information gains. (B) Along with *Peptococcaceae.g_rc4.4* and *Christensenellaceae*, *Bacteroidales.f__S24.7* and *Erysipelotrichaceae* had high ReliefF weights. (C,D) Two PCAs using *Bacteroidales.f__S24.7*, one with *Ruminococcaceae* and the other with *Clostridiales*, were conducted using R. While their first principal components did not change with tumor, their second ones did (Wilcoxon $p < 0.05$, T-test $p < 0.05$ (D)).

2.4. Prediction of Tumor Presence Using Microbial Taxa Involved in Altered Co-Occurrences

Since the second principal components involving *Bacteroidales.f__S24.7*, *Ruminococcaceae*, and *Clostridiales* were found to significantly differ with tumor presence, the proportions of those taxa, along with those of *Peptococcaceae.g_rc4.4*, *Christensenellaceae*, and *Erysipelotrichaceae*, were implemented as a mouse melanoma signature (Figure 4A,B). The 10-fold stratified random sampling used to obtain melanoma prediction results with machine-learning algorithms was performed by randomly selecting 90% of the mouse samples to train the algorithms and then testing them with the remaining 10% of samples (Figure 4A). This process was repeated 10 times, and the prediction results were averaged over those repeats (Figure 4A). Using this protocol, the highest percent accuracy in melanoma prediction was achieved by the neural network, with 80% (Figure 4A,B). Thus, the implementation of microbial taxa indicated by the second principal components in the prediction signature allowed for the identification of melanoma presence.



(A)

Figure 4. Cont.

10-fold Stratified Random Sampling with Taxon Proportions

| <i>Model</i> | <i>AUC</i> | <i>CA</i> | <i>F1</i> | <i>Precision</i> | <i>Recall</i> |
|---------------------|------------|-----------|-----------|------------------|---------------|
| Neural Network | 0.792 | 0.800 | 0.794 | 0.838 | 0.800 |
| SVM | 0.771 | 0.783 | 0.773 | 0.849 | 0.783 |
| AdaBoost | 0.750 | 0.750 | 0.747 | 0.764 | 0.750 |
| CN2 rule inducer | 0.775 | 0.733 | 0.733 | 0.734 | 0.733 |
| Random Forest | 0.861 | 0.733 | 0.729 | 0.751 | 0.733 |
| kNN | 0.660 | 0.667 | 0.665 | 0.670 | 0.667 |
| Naive Bayes | 0.852 | 0.667 | 0.661 | 0.679 | 0.667 |
| Tree | 0.686 | 0.650 | 0.645 | 0.659 | 0.650 |
| Logistic Regression | 0.590 | 0.550 | 0.436 | 0.763 | 0.550 |

(B)

Figure 4. Implementation of microbial taxa implicated in second principal components accurately predict tumor presence. (A) Using Orange3, 10-fold stratified shuffle splits were performed. (B) Using a prediction signature which included *Bacteroidales.f__S24.7*, *Ruminococcaceae*, and *Clostridiales*, implicated in the second principal components, resulted in an average accuracy of 80% achieved with a Neural Network classifier. AUC, area under the curve; CA, classification accuracy; F1, F1 score).

3. Discussion

Our findings demonstrated that the presence of a mouse melanoma tumor can be detected through the altered gut microbial proportions using classification algorithms. By using the gut microbial taxa to model tumor presence, it became apparent that such a condition manifested in more ways than just changes in individual amounts of certain taxa. Indeed, one of the main implications of this study is that considering gut microbial taxa co-occurrences and dependencies in predictive modeling can significantly increase predictive power in melanoma, more so than analyzing only statistical significance between groups. This concept of intertaxa correlations in modeling microbial-based conditions has wide applications in the interpretation of the gut microbiota, as it suggests that the role of an individual taxon in manifesting a biological phenotype is not solely attributed to its unique characteristics [17,18]. Rather, this role also depends on the extent to which a single taxon can communicate and affect other taxa through various mechanisms, from quorum sensing to metabolite production [20–23].

Despite this apparent, predictive relationship between murine melanoma and the gut microbiota, certain experimental limitations still existed. The primary limitation for consideration was the external validity of these results. It is often the case that gut microbiota data do not directly correspond between murine and human subjects, with various mechanisms implicated, from general differences in GI physiology to lifestyle, epigenetics, and immune responses [24–26]. Thus, in order for gut microbial associations to be implemented in clinical cancer diagnoses, further work needs to be done to elucidate pertinent taxa in a variety of human populations and pathophysiological states, including cancer, as well as the interaction between shifts in gut microbial content and certain factors such as diet and lifestyle. Most pertinent to patient treatment is the level of interaction between host immune responses and the gut microbiota, as antitumor immunity and immunotherapies may affect prediction outcomes [27,28]. These studies would also need to consider the correlation between patient stool sampling and gut microbial content with cancer presence, as sampling variation may be a confound [24]. Finally, since our gut microbiota data had a

certain level of variation, other parameters should be considered in the future predictive modeling of human melanoma, such as biochemical and clinical observations [29].

In the statistical analysis of gut microbial taxa, algorithms have been developed to accurately detect the presence of these intertaxa co-occurrences [30–32]. Such algorithms for the detection of microbial “co-occurrence networks” include Sparse Inverse Covariance Estimation for Ecological Association Inference (SPEIC-EASI) and Sparse Correlations for Compositional Data (SparCC) [31–33]. However, despite these advances in the statistical detection of these interactions, there has not been as much work to determine their efficacy in different types of classification algorithms in conditions such as melanoma. In fact, their presence in predictive models has generally been discouraged, as the collinearity they create have been shown to compromise the performance of many model types [34–36]. Further, even for models that can more readily account for collinearity, the use of such interactions in these models does not consistently increase the performance of those models [34–36]. Thus, there is a necessity for a new statistical interpretation of intertaxa co-occurrences in order for them to be optimally utilized in a predictive model. Perhaps new insights into such interpretations can be eventually made when taxa indicated by shifts in co-occurrence networks are further tested in more architecturally complex algorithms such as deep-learning neural networks.

Traditionally, one of the most common procedures in dealing with collinearity between variables such as microbial taxa is the use of principal components in principal component analysis (PCA) [34–37]. By definition, the resulting principal components do not significantly correlate with each other, and are thus used in various model types [34–37]. These components are not usually interpretable from the perspective of the original data because they are linear transformations of that data [34–37]. However, if a small number of variables (e.g., two or three) is used, the principal components can be more easily interpreted [34–37]. In this study, PCA analysis was able to differentiate the two groups of mice successfully; however, much work still needs to be done to characterize the significance of individual PCs in different situations, such as in other clinically relevant tumor types.

4. Methods

4.1. Cell Culture

B16-F10 cells (ATCC) were cultured in RPMI 1640 plus 10% heat-inactivated fetal bovine serum (Atlanta Biologicals, Flowery Branch, GA, USA), 2 mM L-glutamine (Mediatech, Manassas, VA, USA), and 1% penicillin/streptomycin (Mediatech).

4.2. Mouse Experiments

C57BL/6 mice (B6; no. 00664; Jackson Laboratory) were housed in a specific pathogen-free facility at the Rutgers Cancer Institute of New Jersey. Experiments involving animals were carried out in accordance with respective Institutional Animal Care and Use Committee (IACUC) and Institutional Biosafety Committee (IBC) guidelines.

In the first experiment, 35 B6 male mice, aged 6 to 8 weeks old from the Jackson Laboratory were intradermally challenged in the right flank with 10^5 cells of the highly aggressive and poorly immunogenic melanoma B16 cell line ($n = 19$) [17] or phosphate buffered saline (PBS) ($n = 16$) under isoflurane anesthesia. Mice were fed regular chow according to animal care institutional guidelines. Fecal sample collection to compare tumor-bearing to non-tumor-bearing mice was carried out on day 10, when tumors were approximately 25–50 mm². Samples were stored immediately at -80 °C until DNA extraction [38] and sequencing.

The second experiment at this facility followed the identical protocol, using 21 B6 male mice aged 6 to 8 weeks old that were intradermally challenged in the right flank with 10^5 cells of the highly aggressive and poorly immunogenic melanoma B16 cell line ($n = 11$) [17] or phosphate buffered saline (PBS) ($n = 10$) under isoflurane anesthesia. Fecal sample collection to compare tumor-bearing to non-tumor-bearing mice was carried out on

day 16, when tumors were approximately 25–50 mm² in diameter. Samples were stored immediately at –80 °C until DNA extraction [38] and sequencing.

4.3. DNA Extraction

Fecal pellets were homogenized and extracted using the QIAamp PowerFecal DNA Extraction kit following the manufacturer's protocols [39].

4.4. 16S rRNA Gene Sequencing and Data Analysis

The 16S rRNA genes were amplified from purified DNA using PCR primers specific to the V3–V4 region of the 16S rRNA gene and sequenced by Illumina MiSeq in a 2 × 150 bp configuration at the Rutgers New Jersey Medical School Genomics Core. Quantitative Insights Into Microbial Ecology (QIIME) software was used for open-reference operational taxonomic unit (OTU) classification with OTU clustering at 0.97, followed by rarefaction and taxonomic classification of de novo OTUs [40].

4.5. qPCR for Bacterial Load and Taxa Assays

Bacterial loads of extracted fecal DNA were determined by qPCR. DNA were quantified against a standard curve, and the results were normalized to the weight of fecal samples [40].

4.6. Taxon Comparisons, Analyses, and Statistical Modeling

Using the R programming language, microbial taxa between tumor-bearing and PBS control mice were compared using Welch's *t*-test as well as the Mann–Whitney *U* test (a *p*-value of <0.05 was considered to denote statistically significant differences). Between these two groups of mice, general taxa and comparison attributes were determined using the Orange3 v3.27.1 data-mining program and the CORElearn package in CRAN. PCA analysis and principal components were determined using the prcomp function in R. General machine-learning model analyses and cross-validation procedures were performed using the Orange3 program with these settings:

The neural network was a 100-neuron single hidden layer that used the ReLu activation function and the Adam solver.

The support vector machine (SVM) used a radial basis function (RBF) kernel with a cost of 1.0 and a regression loss epsilon of 0.1.

The AdaBoost used a SAMME.R classification algorithm with a linear regression loss function, 50 estimators, and learning rate of 1.0.

The CN2 rule inducer used entropy as the evaluation measure, a beam width of 5, and a maximum rule length of 5.

The random forest used a 12-tree ensemble with subsets split no smaller than 5.

The k-nearest neighbor (kNN) used 5 neighbors and considered the Euclidean distance and uniform weights.

For the naïve Bayes, the attributes were not weighted.

Tree used a maximal tree depth of 100 and subsets not split smaller than 5.

In the logistic regression, a ridge regularization was implemented.

Quality parameters for this model were determined using an internal 10-fold stratified shuffle split, with 90% of the samples selected for training and the remaining 10% for testing in Orange3. Results were graphed using the ggplot2, ggrepel, and ggpubr packages in CRAN, as well as Orange3 and Cytoscape v3.7.2. Heatmaps were generated using the ComplexHeatmap package in CRAN. Tables were formatted using the sjPlot package in CRAN.

Author Contributions: Conceptualization, J.A.G.-P. and A.Z.; formal analysis, M.R.; investigation, S.M.A., F.J.K., J.H.N., A.L., R.J.P., S.M.D., N.L.H. and R.D.; resources, T.M.K. and J.R.; data curation, M.R.; writing—original draft preparation, M.R.; writing—review and editing, S.M.A., F.J.K., J.H.N., A.L., R.J.P., S.M.D., N.L.H., R.D., T.M.K., J.R., J.A.G.-P. and A.Z.; supervision, J.A.G.-P. and A.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data supporting the findings of this study are available from the corresponding authors upon reasonable request.

Acknowledgments: We would like to thank those at the Rutgers Robert Wood Johnson Medical School, Rutgers New Jersey Medical School, and Rutgers Cancer Institute of New Jersey who offered technical assistance. We would also like to thank those at Rush University Medical Center who provided input and support in the writing of this manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lee, Y.K.; Mazmanian, S.K. Has the microbiota played a critical role in the evolution of the adaptive immune system? *Science* **2010**, *330*, 1768–1773. [[CrossRef](#)] [[PubMed](#)]
2. Faust, K.; Raes, J. Microbial interactions: From networks to models. *Nat. Rev. Genet.* **2012**, *10*, 538–550. [[CrossRef](#)] [[PubMed](#)]
3. Ai, D.; Pan, H.; Li, X.; Gao, Y.; Liu, G.; Xia, L.C. Identifying gut microbiota associated with colorectal cancer using a zero-inflated lognormal model. *Front. Microbiol.* **2019**, *10*, 826. [[CrossRef](#)] [[PubMed](#)]
4. Hooper, L.V.; Littman, D.R.; MacPherson, A.J. Interactions between the microbiota and the immune system. *Science* **2012**, *336*, 1268–1273. [[CrossRef](#)]
5. Wu, L.; Estrada, O.; Zaborina, O.; Bains, M.; Shen, L.; Kohler, J.E.; Patel, N.; Musch, M.W.; Chang, E.B.; Fu, Y.-X.; et al. Recognition of host immune activation by *Pseudomonas aeruginosa*. *Science* **2005**, *309*, 774–777. [[CrossRef](#)]
6. Schwabe, R.F.; Jobin, C. The microbiome and cancer. *Nat. Rev. Cancer* **2013**, *13*, 800–812. [[CrossRef](#)]
7. Geng, J.; Song, Q.; Tang, X.; Liang, X.; Fan, H.; Peng, H.; Guo, Q.; Zhang, Z. Co-occurrence of driver and passenger bacteria in human colorectal cancer. *Gut Pathog.* **2014**, *6*, 26. [[CrossRef](#)]
8. Ruiz-Perez, D.; Guan, H.; Madhivanan, P.; Mathee, K.; Narasimhan, G. So you think you can PLS-DA? *BMC Bioinform.* **2020**, *21*, 2. [[CrossRef](#)]
9. Miller, D.D. The medical AI insurgency: What physicians must know about data to practice with intelligent machines. *NPJ Digit. Med.* **2019**, *2*, 62. [[CrossRef](#)]
10. Chou, I.-C.; Voit, E.O. Recent developments in parameter estimation and structure identification of biochemical and genomic systems. *Math. Biosci.* **2009**, *219*, 57–83. [[CrossRef](#)]
11. Saeys, Y.; Inza, I.; Larrañaga, P. A review of feature selection techniques in bioinformatics. *Bioinformatics* **2007**, *23*, 2507–2517. [[CrossRef](#)] [[PubMed](#)]
12. Half, E.; Keren, N.; Reshef, L.; Dorfman, T.; Lachter, I.; Kluger, Y.; Reshef, N.; Knobler, H.; Maor, Y.; Stein, A.; et al. Fecal microbiome signatures of pancreatic cancer patients. *Sci. Rep.* **2019**, *9*, 16081. [[CrossRef](#)] [[PubMed](#)]
13. Ringnér, M. What is principal component analysis? *Nat. Biotechnol.* **2008**, *26*, 303–304. [[CrossRef](#)] [[PubMed](#)]
14. Liu, W.-T.; Marsh, T.; Cheng, H.; Forney, L.J. Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA. *Appl. Environ. Microbiol.* **1997**, *63*, 4516–4522. [[CrossRef](#)] [[PubMed](#)]
15. Lu, M.; Fan, Z.; Xu, B.; Chen, L.; Zheng, X.; Li, J.; Znati, T.; Mi, Q.; Jiang, J. Using machine learning to predict ovarian cancer. *Int. J. Med. Inform.* **2020**, *141*, 104195. [[CrossRef](#)] [[PubMed](#)]
16. Osborn, A.M.; Moore, E.R.B.; Timmis, K.N. An evaluation of terminal-restriction fragment length polymorphism (T-RFLP) analysis for the study of microbial community structure and dynamics. *Environ. Microbiol.* **2000**, *2*, 39–50. [[CrossRef](#)]
17. Sivan, A.; Corrales, L.; Hubert, N.; Williams, J.B.; Aquino-Michaels, K.; Earley, Z.M.; Benyamin, F.W.; Lei, Y.M.; Jabri, B.; Alegre, M.-L.; et al. Commensal bifidobacterium promotes antitumor immunity and facilitates anti-PD-L1 efficacy. *Science* **2015**, *350*, 1084–1089. [[CrossRef](#)]
18. Derrien, M.; Belzer, C.; de Vos, W.M. *Akkermansia muciniphila* and its role in regulating host functions. *Microb. Pathog.* **2017**, *106*, 171–181. [[CrossRef](#)]
19. Zeng, H.; Ishaq, S.; Liu, Z.; Bukowski, M. Colonic aberrant crypt formation accompanies an increase of opportunistic pathogenic bacteria in C57BL/6 mice fed a high-fat diet. *J. Nutr. Biochem.* **2018**, *54*, 18–27. [[CrossRef](#)]
20. Miller, M.B.; Bassler, B.L. Quorum sensing in bacteria. *Annu. Rev. Microbiol.* **2001**, *55*, 165–199. [[CrossRef](#)]
21. Mulcahy, L.R.; Isabella, V.M.; Lewis, K. *Pseudomonas aeruginosa* biofilms in disease. *Microb. Ecol.* **2014**, *68*, 1–12. [[CrossRef](#)] [[PubMed](#)]
22. Li, F.; Cimdins, A.; Rohde, M.; Jansch, L.; Kaever, V.; Nimtz, M.; Römling, U. DncV synthesizes cyclic GMP-AMP and regulates biofilm formation and motility in *Escherichia coli* ECOR31. *mBio* **2019**, *10*, e02492-18. [[CrossRef](#)] [[PubMed](#)]
23. Chen, L.; Li, X.; Zhou, X.; Zeng, J.; Ren, Z.; Lei, L.; Kang, D.; Zhang, K.; Zou, J.; Li, Y. Inhibition of *Enterococcus faecalis* growth and biofilm formation by molecule targeting cyclic di-AMP synthetase activity. *J. Endod.* **2018**, *44*, 1381–1388.e2. [[CrossRef](#)] [[PubMed](#)]

24. Zmora, N.; Zilberman-Schapira, G.; Suez, J.; Mor, U.; Dori-Bachash, M.; Bashirdes, S.; Kotler, E.; Zur, M.; Regev-Lehavi, D.; Brik, R.B.-Z.; et al. Personalized gut mucosal colonization resistance to empiric probiotics is associated with unique host and microbiome features. *Cell* **2018**, *174*, 1388–1405.e21. [[CrossRef](#)] [[PubMed](#)]
25. Thevaranjan, N.; Puchta, A.; Schulz, C.; Naidoo, A.; Szamosi, J.; Verschoor, C.P.; Loukov, D.; Schenck, L.P.; Jury, J.; Foley, K.P.; et al. Age-associated microbial dysbiosis promotes intestinal permeability, systemic inflammation, and macrophage dysfunction. *Cell Host Microbe* **2017**, *21*, 455–466.e4. [[CrossRef](#)]
26. Gibbons, S.M.; Kearney, S.M.; Smillie, C.S.; Alm, E.J. Two dynamic regimes in the human gut microbiome. *PLoS Comput. Biol.* **2017**, *13*, e1005364. [[CrossRef](#)]
27. Wong, S.H.; Zhao, L.; Zhang, X.; Nakatsu, G.; Han, J.; Xu, W.; Xiao, X.; Kwong, T.N.Y.; Tsoi, H.; Wu, W.K.K.; et al. Gavage of fecal samples from patients with colorectal cancer promotes intestinal carcinogenesis in germ-free and conventional mice. *Gastroenterology* **2017**, *153*, 1621–1633.e6. [[CrossRef](#)]
28. Elkrief, A.; Derosa, L.; Zitvogel, L.; Kroemer, G.; Routy, B. The intimate relationship between gut microbiota and cancer immunotherapy. *Gut Microbes* **2019**, *10*, 424–428. [[CrossRef](#)]
29. McDonald, D.; Hyde, E.; Debelius, J.W.; Morton, J.T.; Gonzalez, A.; Ackermann, G.; Aksenov, A.A.; Behsaz, B.; Brennan, C.; Chen, Y.; et al. American gut: An open platform for citizen science microbiome research. *mSystems* **2018**, *3*, e00031-18. [[CrossRef](#)]
30. San-Juan-Vergara, H.; Zurek, E.; Ajami, N.J.; Mogollon, C.; Peña, M.; Portnoy, I.; Velez, J.; Cadena-Cruz, C.; Diaz-Olmos, Y.; Hurtado-Gómez, L.; et al. A Lachnospiraceae-dominated bacterial signature in the fecal microbiota of HIV-infected individuals from Colombia, South America. *Sci. Rep.* **2018**, *8*, 4479. [[CrossRef](#)]
31. Kurtz, Z.D.; Mueller, C.L.; Miraldi, E.R.; Littman, D.R.; Blaser, M.J.; Bonneau, R.A. Sparse and compositionally robust inference of microbial ecological networks. *PLoS Comput. Biol.* **2015**, *11*, e1004226. [[CrossRef](#)] [[PubMed](#)]
32. Kim, S.; Thapa, I.; Zhang, L.; Ali, H. A novel graph theoretical approach for modeling microbiomes and inferring microbial ecological relationships. *BMC Genom.* **2019**, *20*, 945. [[CrossRef](#)] [[PubMed](#)]
33. Hirano, H.; Takemoto, K. Difficulty in inferring microbial community structure based on co-occurrence network approaches. *BMC Bioinform.* **2019**, *20*, 329. [[CrossRef](#)]
34. De Marco, P.J.; Nobrega, C.C. Evaluating collinearity effects on species distribution models: An approach based on virtual species simulation. *PLoS ONE* **2018**, *13*, e0202403. [[CrossRef](#)] [[PubMed](#)]
35. Pasanen, L.; Holmström, L.; Sillanpää, M.J. Bayesian LASSO, scale space and decision making in association genetics. *PLoS ONE* **2015**, *10*, e0120017. [[CrossRef](#)]
36. Schisterman, E.F.; Perkins, N.J.; Mumford, S.L.; Ahrens, K.A.; Mitchell, E.M. Collinearity and causal diagrams: A lesson on the importance of model specification. *Epidemiology* **2017**, *28*, 47–53. [[CrossRef](#)]
37. Jiang, M.; Zhu, L.; Wang, Y.; Xia, L.; Shou, G.; Liu, F.; Crozier, S. Application of kernel principal component analysis and support vector regression for reconstruction of cardiac transmembrane potentials. *Phys. Med. Biol.* **2011**, *56*, 1727–1742. [[CrossRef](#)]
38. Matson, V.; Fessler, J.; Bao, R.; Chongsuwat, T.; Zha, Y.; Alegre, M.-L.; Luke, J.J.; Gajewski, T.F. The commensal microbiome is associated with anti-PD-1 efficacy in metastatic melanoma patients. *Science* **2018**, *359*, 104–108. [[CrossRef](#)]
39. McGaughey, K.D.; Yilmaz-Swenson, T.; Elsayed, N.; Cruz, D.A.; Rodriguez, R.R.; Kritzer, M.D.; Peterchev, A.V.; Gray, M.; Lewis, S.; Roach, J.; et al. Comparative evaluation of a new magnetic bead-based DNA extraction method from fecal samples for downstream next-generation 16S rRNA gene sequencing. *PLoS ONE* **2018**, *13*, e0202858. [[CrossRef](#)]
40. Lei, Y.M.; Chen, L.; Wang, Y.; Stefka, A.; Molinero, L.L.; Theriault, B.; Aquino-Michaels, K.; Sivan, A.S.; Nagler, C.R.; Gajewski, T.F.; et al. The composition of the microbiota modulates allograft rejection. *J. Clin. Investig.* **2016**, *126*, 2736–2744. [[CrossRef](#)]