ELUCIDATING GUT MICROBIOTA-IMMUNE INTERACTIONS IN AN ANIMAL MODEL OF ANKYLOSING SPONDYLITIS

By

Mark T. Klick

A THESIS

Presented to the Department of Biomedical Informatics and Computational Biology and the Oregon Health & Science University School of Medicine in partial fulfillment of the requirements for the degree of

Master of Science

September 2017

School of Medicine

Oregon Health & Science University

Certificate of Approval

This is to certify that the Master's Thesis of

Mark T. Klick

"Elucidating Gut Microbiota-Immune Interaction in an Animal Model of Ankylosing Spondyltis"

Has been approved

Abstract

Introduction: Ankylosing spondylitis (AS) is a chronic inflammatory disorder in which aberrant immune responses to the intestinal microbiota are thought to drive pathogenesis. The HLA-B27 transgenic rat is a foremost translational model of disease, with rats expressing the major human risk allele for AS. We have shown previously the IgA response to gut microbes is strongly elevated in this model, but the specificity of this response remains unclear. Here we used the novel IgA-SEQ technique to test the hypothesis that HLA-B27 expression alters the microbial repertoire of the intestinal IgA response.

Methods: Feces were collected from 40 adult HLA-B27 transgenic rats (TG) and 32 wild-type (WT) controls. IgA-coated and uncoated fecal bacteria were sorted by flow cytometry and these fractions were subjected to 16s rRNA gene sequencing (IgA-SEQ). Relative abundance of bacteria in these fractions was determined using QIIME and enrichment of bacteria amongst the IgA positive fraction was compared between transgenic and control rats. To test the hypothesis that HLA-B27 expression alters the microbial repertoire of the intestinal IgA response we compared the IgA response of control and transgenic rats by quantifying the IgA response via two metrics: IgA Coating Index (ICI Score) and IgA Index. We compared the IgA response of control and transgenic rats, as quantified by the two metrics, using two approaches: traditional and predictive modeling. These two approaches were validated/replicated in a smaller human dataset of 12 HLA-B27 positive and 12 HLA-B27 negative fecal samples collected during routine colonoscopies for colon cancer screening.

Results: Our analysis identified several bacterial taxa that have a significantly enriched IgA response in the HLA-B27 rats compared to healthy controls. We also found that in the HLA-B27 rat dataset the IgA Index was a more robust IgA response quantification metric than the ICI Score. In our human HLA-B27 dataset we found that some of the highly IgA coated bacteria in the HLA-B27 rats were also exhibiting similar IgA coating in the HLA-B27 humans. We also found several other bacterial taxa, unique to the human dataset, that have a significantly enriched IgA response in the HLA-B27 humans compared to healthy controls.

Discussion: IgA-SEQ revealed HLA-B27 expression dramatically altered the targeting of gut microbes by the intestinal IgA response. We have elucidated a gut microbiota-immune interaction phenotype in the HLA-B27 rat model and gained valuable insight into the immunopathogenesis of AS. We have also validated/replicated our approach in the HLA-B27 humans identifying some similarities to the rat model and other interesting unique-to-human IgA-coated bacterial taxa to study further. This technique could also be used to better understand microbiota-immune interactions in other chronic inflammatory diseases, such as AS, where dysregulated immune responses to gut bacteria are thought to be a key pathophysiological mechanism.

Keywords: Ankylosing Spondylitis, IgA-SEQ, microbiome, Bioinformatics

Background

1.1 The gut microbiota and disease

Evidence from both animal models and translational human studies have established that changes in the composition of the gut microbiome influence and can drive the development of inflammatory, autoimmune, metabolic, and non inflammatory diseases.[4,10] These perturbations from the 'normal' gut microbiota of healthy individuals observed in disease are referred to as dysbiosis. Indeed, dysbiosis has been widely reported in the inflammatory bowel disease (IBD) research field, which includes Crohn's disease and ulcerative colitis, where specific members of the gut microbiota have been identified as disease-causing 'pathobionts' in several transgenic immuno-compromised mouse models of colitis.[2,10] Subsequently, the identification of pathobionts has extended beyond the IBD field to other diseases less obviously linked to the intestine, for instance age-related macular degeneration, neurological disorders like Parkinson's disease, and maybe even urgency urinary incontinence.[22,23,24]

While community-level profiling has helped identify several of these 'pathobionts', these are typically microbes that are present in one population and absent in another, or microbes that are vastly over-represented in the diseased population compared to the healthy population. A significant limitation of the conventional community-wide 16S profiling approach is that it does not account for microbes that may exhibit equivalent colonization in two populations yet elicit a different host response. Specifically, the host immune response may be critically relevant to inflammatory disease and to our understanding of its immunopathogenesis.

In the gut, there are more immune cells (T-cells, B-cells) and many immune mediators, such as immunoglobulins, than anywhere else in the body.[4] The gut's microbial communities are constantly being monitored by the host's intestinal immune system and IgA performs most of that monitoring.[4] Some commensal bacteria in the gut, such as Segmented Filamentous Bacteria (SFB) are known to interact with the intestinal immune system and induce T-cell responses that cause systemic inflammation.[4,6] Robust activation of intestinal T-cell responses is also known to activate B cells to produce secretory IgA.

The gut is a powerful regulator of the host's immune system because it helps to maintain the delicate balance between inflammation and regulation.[4,10] IgA regulates the composition of the gut microbiota and prevents harmful/unnecessary immune system responses by coating specific strains of potentially harmful microbiota which prevents those strain's direct interaction with the host's epithelium.[4] Bacteria, like SFB, that adhere to the gut (intestinal) epithelium come into direct contact with the host and elicit more specificity from IgA.[2,4,10] Furthermore, bacteria that elicit more specificity from IgA, or are more highly IgA coated, are generally considered more pathogenic.[4,6]

1.2 IgA-SEQ

The specific strains of bacteria that are highly bound by IgA can be identified using a novel method, developed by Flavell and colleagues called IgA SEQ.[6] This technique builds upon merely providing a picture of overall bacterial abundance, adding identification of microbes that are specifically enriched for IgA coating, indicating bacteria that are highly targeted by the immune response. In order to do this, fecal bacteria are stained with a fluorescently-conjugated

anti-host IgA antibody (Fig 1). This staining permits IgA-coated bacteria to be separated from their uncoated counterparts by a fluorescently activated cell sorting (FACS) instrument on the basis of differential fluorescence. The isolated IgA positive and IgA negative fractions can then be sequenced to ascertain the microbial composition of these populations by DNA extraction and sequencing of the 16s ribosomal rRNA gene.

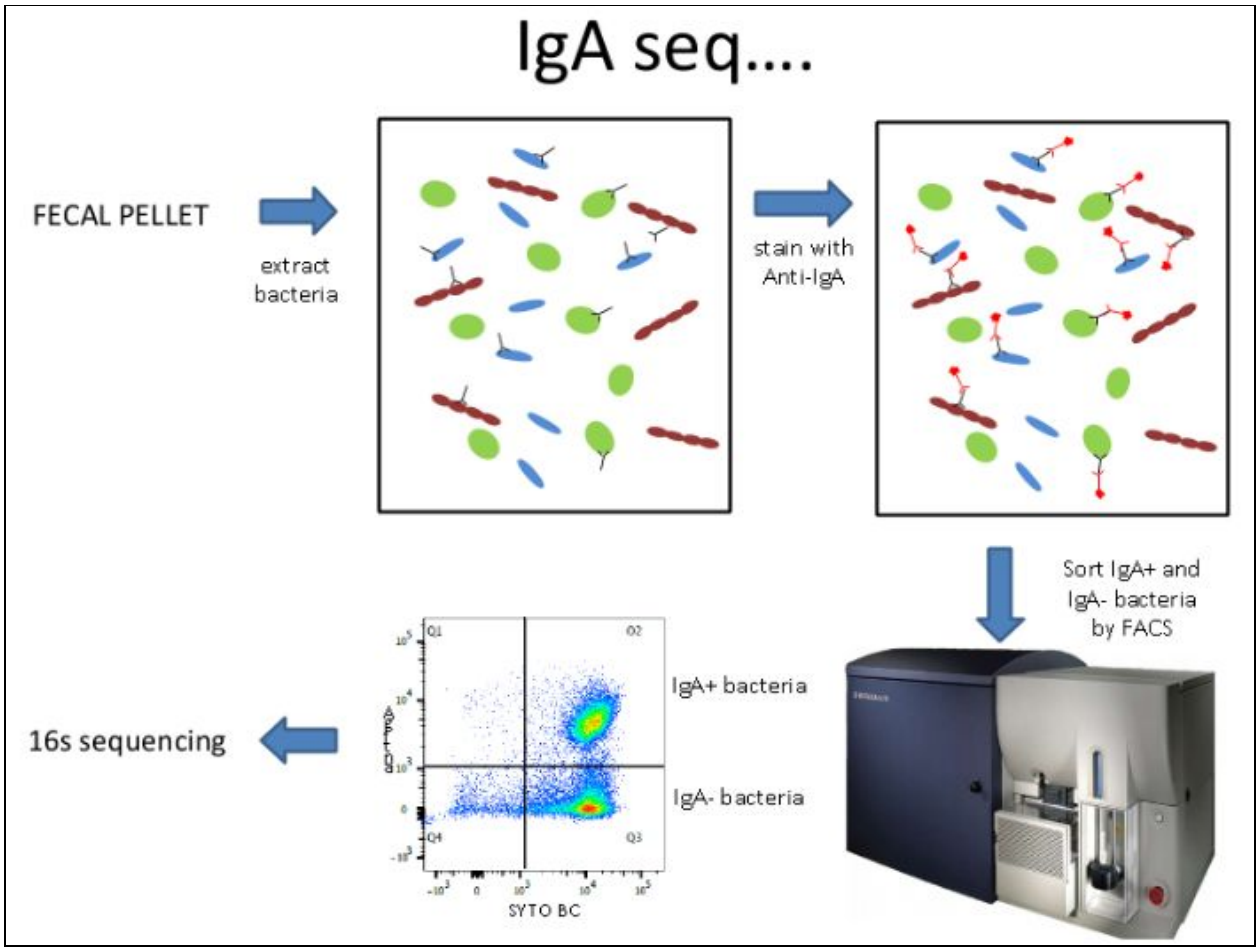


Figure 1. The methodology of IgA-SEQ. Feces were collected from adult HLA-B27 transgenic rats and WT controls (n=32-40 per group). IgA-coated and uncoated fecal bacteria were sorted by flow cytometry and these fractions were subjected to 16s rRNA gene sequencing (IgA-SEQ). These 16s sequencing reads are grouped into OTUs and taxonomy is assigned via QIIME then this data is imported into R for further analysis [14,20,21].

Like conventional community-wide profiling studies of the entire microbial community of the gut, IgA-SEQ also utilizes 16s rRNA gene sequencing to determine microbial composition of a given sample. To do this, DNA is extracted from a microbial sample and the 16s rRNA gene is amplified using primers which target conserved regions of the 16s rRNA gene found in all prokaryotes. The sequenced 16S rRNA gene reads can be identified as belonging to specific bacterial species by aligning the reads to known reference sequences found in databases like SILVA[27] or GREENGENES[28]. The key difference between IgA seq and conventional community-wide 16S profiling is that the bacteria is first sorted into two fractions using fluorescent activated cell sorting (FACS). FACS allows for the 16S sequencing of each of the two fractions, i.e., the fraction coated by IgA and the fraction not coated by IgA.

IgA-SEQ has been used successfully to identify potentially pathogenic/colitogenic bacteria in IBD and malnutrition.[2,5,18] For instance, Palm et al found highly IgA coated microbes induced an inflammatory phenotype and increased colitis severity upon adoptive transfer to murine recipients in a mouse model of IBD.[2] IgA-SEQ has also been used to monitor and describe the development of mucosal IgA responses showing that IgA-SEQ can be used to "define gut mucosal immune development in health and disease states".[3] The successful uses of IgA-SEQ in these studies[2,3,5,11] illustrate that IgA-SEQ is a powerful technique for identifying bacterial taxa that provide a strong stimulus to the host's immune system, i.e., provides scientists the means to identify immunogenic/pathogenic bionts in their experimental investigations.[4,6]

Identifying which biota are potentially pathogenic in a disease like AS and many other inflammatory diseases using IgA-SEQ is of great importance. Performing analysis on 16S data is already challenging and the increased complexity of IgA-seq data, i.e., the different IgA coated fractions, makes this more challenging. Determining what manipulation of the data and statistical approaches are appropriate is not a trivial task, thus we have provided a best practice IgA-SEQ workflow implemented in R which can be used on other data sets (see https://github.com/markklick206/IgA-SEQ_Workflow).[14,17,20,21]

1.3 Ankylosing Spondylitis

Ankylosing spondylitis (AS) and other spondyloarthropathies are hugely debilitating diseases with significant unmet clinical need. These diseases typically manifest themselves at 20-30 years of age with chronic inflammation of the spine and sacroiliac joints. AS alone affects roughly 1% of the U.S. population, representing some 40,000 Oregonians with an estimated annual economic burden of $15-20,000 per patient in healthcare costs and lost productivity. Like IBD and other chronic inflammatory diseases, there are several observations that indicate the intestinal microbiota may be critical for the pathogenesis of IBD.[19] These include:

1. AS patients exhibit a dysbiotic gut phenotype relative to healthy controls.[29]
2. Over half of AS patients exhibit signs of subclinical bowel inflammation implicating disease overlap with IBD.[30]
3. Germ-free animals that are sterile and hence lack an intestinal microbiota do not develop AS-like symptoms.[31]

HLA-B27 is a major risk factor for ankylosing spondylitis, acute anterior uveitis, reactive arthritis and other spondyloarthropathies. Indeed, some 90% of AS patients are HLA-B27 positive versus a prevalence of 5-10% in the general population. Moreover, transgenic rats that express human HLA-B27 go on to develop penetrant HLA-B27-associated inflammatory sequelae such as bowel and joint disease.[32] This transgenic model is an excellent model of disease, with animals developing a hyper-active immune response to the gut microbiota, accompanied by large alterations to the gut microbiota.[1]

The Asquith lab has shown previously that the IgA response to the intestinal microbiota is dramatically increased in HLA-B27 transgenic (TG) rats vs healthy controls.[1] The identity of bacteria targeted by this IgA response are unknown. However, IgA-SEQ has the ability to determine which bacteria may be differentially targeted by the IgA response in this transgenic model. This study consists of 32 WT and 40 TG rats each sorted into IgA-negative (IgA-neg),

IgA-positive (IgA-pos) and All-Bacteria(AllBac) fractions and subjected to 16s rRNA sequencing..

These animals are a robust model of AS. The HLA-B27 rat model could be considered the foremost translational model of AS because animals manifest with bowel inflammation at 8-10 weeks of age.[1] We already mentioned our interest in comparing the IgA response between genotype, but our HLA-B27 dataset also consists of rats of different age groups. If the HLA-B27 32 WT and 40 TG rats develop bowel inflammation at 8-10 weeks old, then it makes sense that 12 weeks old is considered an early time point in disease progression/development and 16 weeks old being considered a later time point with established disease. This later time point is also when animals start to manifest with extra-intestinal symptoms, namely arthritis.[1]

We have a smaller human dataset of HLA-B27 positive and HLA-B27 negative fecal samples collected during routine colonoscopies (n=12-15 per group) in which we will also test the hypothesis that HLA-B27 expression alters the microbial repertoire of the intestinal IgA response using our IgA-SEQ workflow.This validation/replication of our analysis approach allows us to compare the IgA-coated bacterial taxa found in the HLA-B27 rat model to the IgA-coated bacterial taxa found in the HLA-B27 Humans. In other words, the translational potential of the HLA-B27 rat model will be examined. We will also get a glimpse of the intestinal IgA response phenotype in humans which provides novel insight into the immunopathogenesis of AS.

Methods

2.1 Quantifying the IgA Response

To test the hypothesis that HLA-B27 expression alters the microbial repertoire of the intestinal IgA response We will be comparing the IgA response of control and transgenic rats using two approaches. The first approach uses traditional non-parametric statistical tests to compare the IgA response, whereas the second approach uses predictive modeling. Due to the non-normal distribution of the OTU relative abundance data, using a standard t-test to determine statistically significant differences between experimental conditions would not be appropriate. Therefore, the Wilcoxon Signed Rank test and the Wilcoxon Rank Sum test are used to determine any significant differences within/between experimental conditions.[2,3,5,6,11,14] Previous studies that used IgA-SEQ quantified the IgA response using two different metrics: ICI score and IgA Index (Fig. 2).

Palm et. al.[2] used a simple ratio, termed IgA Coating Index (ICI score), of the positive fraction divided by the negative fraction (Fig. 2). A bacterial taxa with an ICI score above one is considered positively IgA enriched, below one is considered negatively IgA enriched, and a ICI score of one is considered not coated. A 0.0002 substitution (detection limit) is made for any 0 relative abundance value in the negative fraction to prevent a division by zero. The authors decided a taxa with an ICI score above 10 was considered highly coated in their experiment.

Kau et. al.[5], Planer et. al.[3], and Viladomiu et. al.[18] used a log normalized ratio, termed the IgA Index (Fig. 2). The IgA Index is bound between a value of negative one and one. A bacterial taxa with a negative IgA Index value is considered negatively IgA enriched, a positive value is considered positively IgA enriched, and a IgA Index of zero is considered not coated. Here, a

substitution of 0.0002 is only made if both negative and positive fractions have 0 relative abundance values.

$$IgACoatingIndex(ICI) = IgA(+)relativeabundance/IgA(-)relativeabundance$$

$$\text{IgA index} = -\frac{\log(\text{IgA}^{+}_{\text{taxon}}) - \log(\text{IgA}^{-}_{\text{taxon}})}{\log(\text{IgA}^{+}_{\text{taxon}}) + \log(\text{IgA}^{-}_{\text{taxon}})}$$

Figure 2. Equations for the two different IgA coating quantification metrics. Both of these metrics were used to quantify the IgA response. Notice the ICI score is unbounded in its range while IgA Index is log normalized so it's bounded between -1 and 1.

2.2 Predictive Modeling Approach

The second approach used to compare the IgA response of control and transgenic rats is to use predictive modeling. This approach allows us to directly compare the different IgA response metrics and determine which metric is more robust (Fig. 2). For the analysis we set up a supervised binary classification problem which assesses the predictive ability of the IgA response in predicting/classifying case vs. control. The goal is to determine what IgA quantification metric, if any, is more predictive by constructing a predictive model to identify what bacteria are important in answering the binary classification problem, i.e., which IgA coated bacteria are best at distinguishing between HLA-B27 vs. control rats.

Using the ICI score and IgA Index matrices (Fig. 3) of the HLA-B27 and control rats as the data for this supervised binary classification problem will not only give insight into what bacterial genera/features are important in predicting/classifying between genotype but also assess the predictive ability/efficacy of the two different IgA response metrics in general. We chose to use a random forest as the algorithm for my predictive model for three reasons. First, a random forest model makes virtually no assumptions on the distribution of the input data. Second, we can assess variable importance in the random forest model to determine what features, or bacterial genera, best discriminate genotype. The third reason is that random forests are commonly used and considered a best practice in the metagenomics field.[3,17,25] To further test the predictive ability of the ICI score and IgA Index we built three other predictive models using three common machine learning algorithms (Fig.3), e.g., Logistic Regression, Linear Discriminant Analysis (LDA), and Classification & Regression Trees (CART).

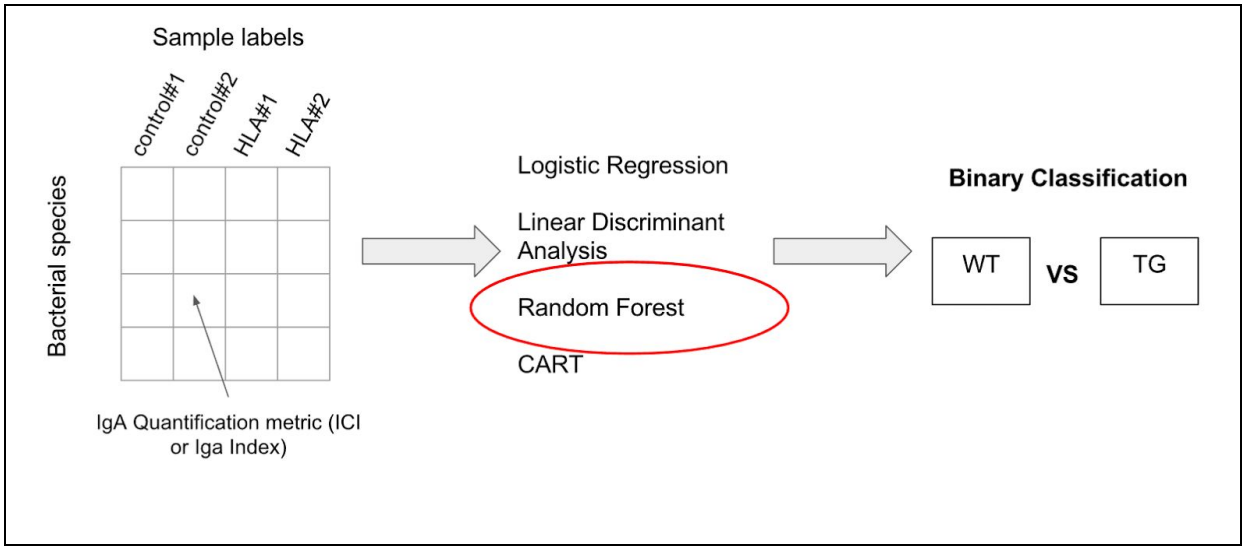


Figure 3. Illustration of the supervised binary classification problem which will address/assess the predictive ability of the IgA response in predicting/classifying HLA-B27 (TG) vs. control rats (WT). Data matrices, various learners, and the binary classification is shown (left to right). Red oval indicates the learner that will be used in this analysis, i.e., a random forest. The other learners will be ran as well.

2.3 Sequence Processing and Taxonomic Identification

16s rRNA gene sequencing was performed with Illumina MiSeq sequencer (Illumina, USA). The sequences were processed using scripts from the workflow package Quantitative Insights into Microbial Ecology (QIIME) version 1.9.0.[15] Individual sequence reads were joined using FASTQ-join (ea-utils, version 1.1.2-537; Aronesty, 2013). The sequencing reads were demultiplexed and sequences were grouped into operational taxonomic units (OTUs) with the open reference approach implemented in pick_open_reference_otus.py, with the uclust algorithm and Green Genes (gg_13_8) reference database. Chimeric sequences were removed with the blast_fragments approach implemented in the identify_chimeric_seqs.py script. The resulting OTU table was imported into R for filtering and statistical analysis.

2.4 IgA-SEQ Workflow

One of the goals of this analysis was to produce a generalized workflow for any arbitrary case vs. control dataset. We use the HLA-B27 rat data set and HLA-B27 human data set as use cases for the IgA-SEQ workflow. A description of the IgA-SEQ workflow used on both HLA-B27 human and rat data sets analysis can be seen in Figure 4.

The IgA-SEQ workflow begins with the key input files for any microbiome study: an OTU table (usually in BIOM format), a phylogenetic tree file, and an experimental metadata mapping file. We use QIIME Python scripts to convert the OTU table into the standardized BIOM format.[12,13] The R package Phyloseq is used to read in the BIOM table and any experimental metadata into the R environment.[20] Exploratory data analysis (EDA) is then performed, e.g., making boxplots to visualize the distributions of the relative abundance data. EDA can identify confounding covariates and allow for careful adjustments based on the EDA such as normalization or removal of outlier samples. One of the most common EDA methods for microbiome data is principal coordinates analysis (PCoA) which allows us to visually identify associations with a

covariate of interest: genotype and age.

After conducting EDA, the relative abundances between the IgA-SEQ fractions are compared via Kruskal-Wallis test comparing the positive and negative IgA fractions, similar to the first step in the widely used tool LEfSe.[7] We also employ a PERMANOVA test on the IgA positive fraction using genotype as our covariate of interest to test for associations between bacterial abundance and genotype. To further supplement our analysis we use predictive modeling to build several random forest models using the log(1+x) normalized relative abundance values for the different IgA-SEQ fractions, i.e., IgA-pos, IgA-neg, and AllBac, to predict genotype and identify discriminative bacteria.[17,25]

After visualizing and performing statistical tests on the relative abundances of the different fractions in each sample, we compute the IgA quantification metrics, i.e., ICI score and IgA Index, for the top 45 most abundant bacterial genera (OTUs) in the AllBac Fraction of the data set. The biological reasoning behind this subsetting choice is that we are interested in microbes/bacterial genera that may exhibit equivalent colonization in control and HLA-B27 subjects yet elicit a different host IgA response. However, we have several other methods for subsetting but the top 45 most abundant bacterial genera in the AllBac Fraction was the most biologically relevant subset. We also look at the distribution of the IgA quantification metrics across the samples and bacterial genera (OTUs) and visualize the differences in the IgA response between control and HLA-B27 via boxplots.

Then we employ our traditional approach which uses non-parametric statistical tests to compare the IgA response between control and HLA-B27 subjects. A Wilcoxon Signed rank test is used within each group, control and HLA-B27, to determine any statistically significant differences in the distribution of the sample's IgA quantification value for each bacterial genera (OTU), i.e., within each genotype we are statistically testing which taxa have an ICI score that $\neq 1$ indicating they are significantly IgA enriched (>1) or IgA unenriched (<1). A Wilcoxon Rank Sum test is used between each group, control and HLA-B27, to determine if there are any significant differences in the distribution of the IgA quantification value for each bacterial genera (OTU's). The results of these statistical comparisons on the IgA-quantification metrics between control and HLA-B27 rats are then visualized via heatmaps and bubble plots.

We then employ our predictive modeling approach by constructing the random forest models, using the ICI and IgA Index, with the random forest implementation in the randomForest R package (**Fig. 3**).[21] Like any standard supervised binary classification we subset our data set into a training and test set, we chose a 80:20 split. We repeat the random forest model building procedure 100 times similar to that performed by Planer et al [3].

Model evaluation metrics, i.e., accuracy, sensitivity and AUC, tell us how well our models using the ICI and IgA Index can discriminate between control and HLA-B27 subjects (how well they did in answering the supervised binary classification problem). We then visualize the two predictive model's variable importance plots which illustrate the bacterial genera that are most important to the random forest, i.e., the bacterial genera that best discriminate between HLA-B27 and control subjects.

To further test the predictive ability of the ICI score and IgA Index we built three other predictive models using three common machine learning algorithms (Fig.3), e.g., Logistic Regression,

Linear Discriminant Analysis (LDA), and Classification & Regression Trees (CART).

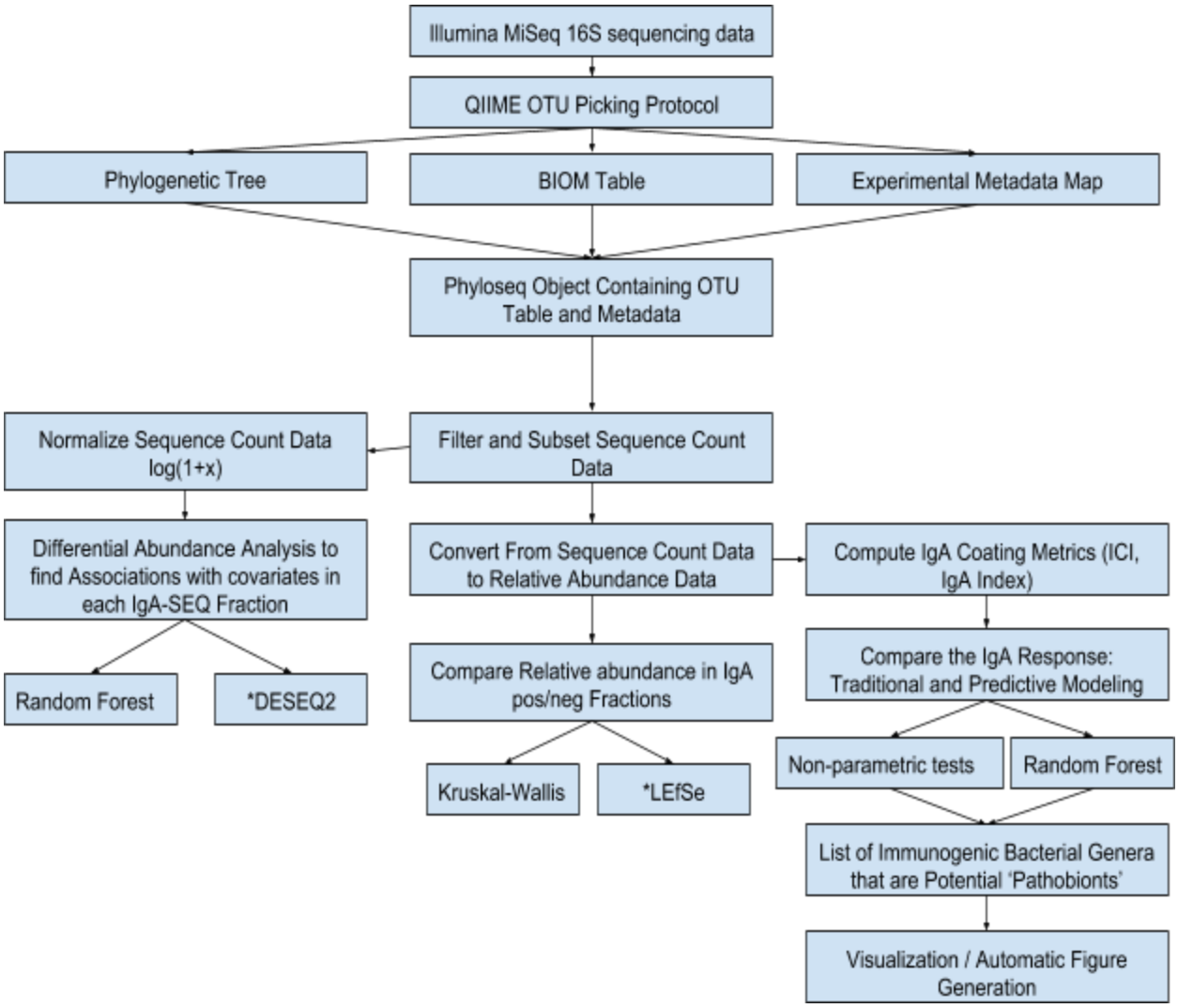


Figure 4. Flow diagram of the IgA-SEQ Workflow used on the HLA-B27 rat dataset and HLA-B27 human dataset. *DESEQ2 was not used and just is shown as an example. *LEfSe was not used and is just shown as an example.

Results

3.1 A Subset of Bacterial Genera are Enriched for IgA Coating in HLA-B27 Rats

There is a clear difference in the relative abundance of bacterial taxa for the various IgA-SEQ fractions: negative, positive, and AllBac in the HLA-B27 rat data set. Like others who have used IgA-SEQ have shown in their datasets, there is also a distinct subset of bacterial genera present in the IgA positive fraction of the HLA-B27 rat model.[2,3,5,18] Figure 5 illustrates the elevated abundance of certain bacterial genera in the IgA positive fraction of the HLA-B27 rats. We also constructed PCoA plots (supplementary materials) for each of the fractions using genotype as

the covariate. These PCoA plots indicate that there is some separation between control and HLA-B27 rats. To further supplement/confirm the differential abundance between genotypes a PERMANOVA test was computed in the IgA positive fraction, and as we expected the result was significant with a p-value<0.02.

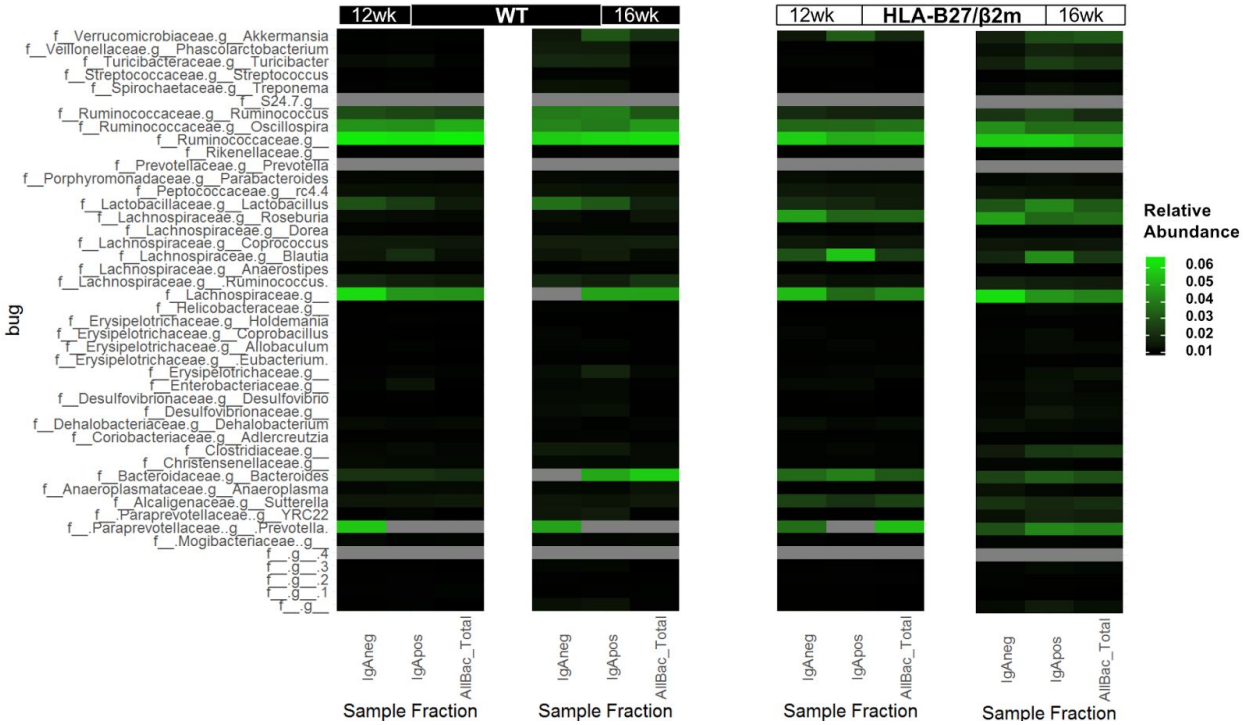


Figure 5. Relative Abundance of IgA-SEQ fractions. Heat maps are used to illustrate the relative abundance values for the top 45 most abundant OTUs in the AllBac Fraction of the HLA-B27 rat data set (microbes that may exhibit equivalent colonization in two populations yet elicit a different host response). The average relative abundance values for IgA-, IgA+ and AllBac fractions are shown for control (WT) and HLA-B27 rats. We are showing the different age groups, 12 and 16 weeks, to visualize the differences in relative abundance in our HLA-B27 16 week old diseased rats. There are several OTUs that have noticeably higher relative abundance in the IgA+ fraction of the HLA-B27 16weeks rats.

Figure 5 shows the differences in relative abundance for each of the IgA-SEQ fractions between the control and HLA-B27 rats. Moreover, the 16 week transgenic rats seem to exhibit a distinct repertoire of bacterial genera that are present in the IgA positive fraction. To further investigate this we employ the two IgA quantification metrics that have been used in literature, e.g., the ICI score and IgA Index. When we look at the IgA response of the top 45 most abundant bacteria in the AllBac fraction quantified by the ICI score and visualized via a heatmap, we can see that HLA-B27 expression dramatically alters the targeting of gut microbes by the intestinal IgA response (Fig. 6). Furthermore, this alteration in the targeting of gut microbes by the intestinal IgA response seems to be more pronounced in the 16 week old HLA-B27 rats. Several bacterial genera standout allowing us to narrow in on a subset of highly IgA coated bacteria that are potentially driving disease in the HLA-B27 rat (Fig. 5-6).

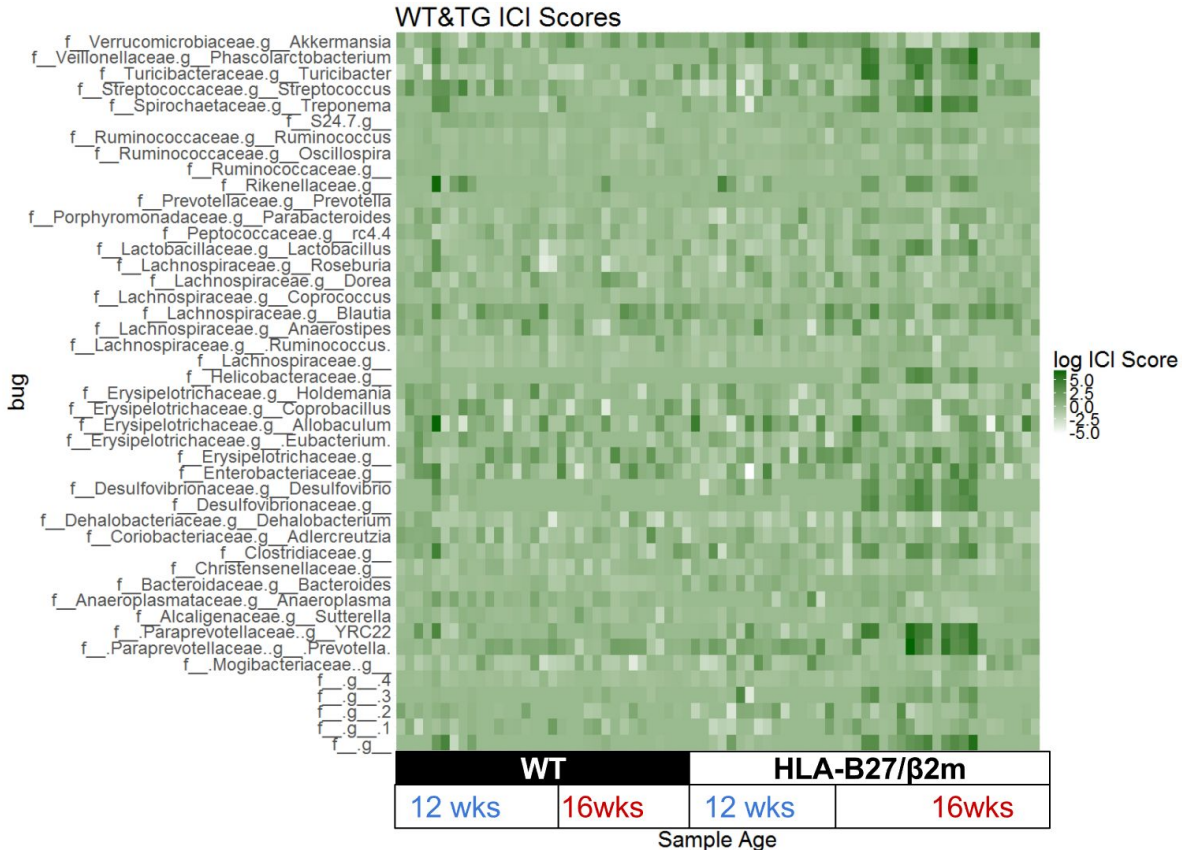


Figure 6. ICI Score. Heat map illustrating the distribution of ICI scores for the top 45 most abundant OTUs in the AllBac Fraction of the HLA-B27 Rat data set (microbes that may exhibit equivalent colonization in two populations yet elicit a different host response). The control (WT) rats are shown on the left subdivided into 12 and 16 week age groups, and the HLA-B27 rats are shown on the right. The IgA response as quantified by the ICI score identifies several OTUs that are 'highly IgA coated' in the HLA-B27 16 weeks samples. *log ICI score is shown here for better visualization.

3.2 HLA-B27 Expression Alters the Microbial Repertoire of the Intestinal IgA Response

Visualizing the ICI score and IgA Index using a boxplot allows us to get a sense of the distribution of the IgA response for each bacterial genus and visually compare the quantified IgA response of control and HLA-B27 rats. Comparing the severity/intensity of the differences in IgA response between control and HLA-B27 rats, we can see that ICI score slightly exaggerates the IgA response and the IgA Index depicts the IgA response less variably. Figure 7 illustrates the fact that no matter how you quantify the IgA response, ICI score or IgA Index, a distinct difference in the IgA repertoire between genotypes can be seen. Moreover, the distribution of IgA-coated microbial taxa identified by these two indices is broadly similar.

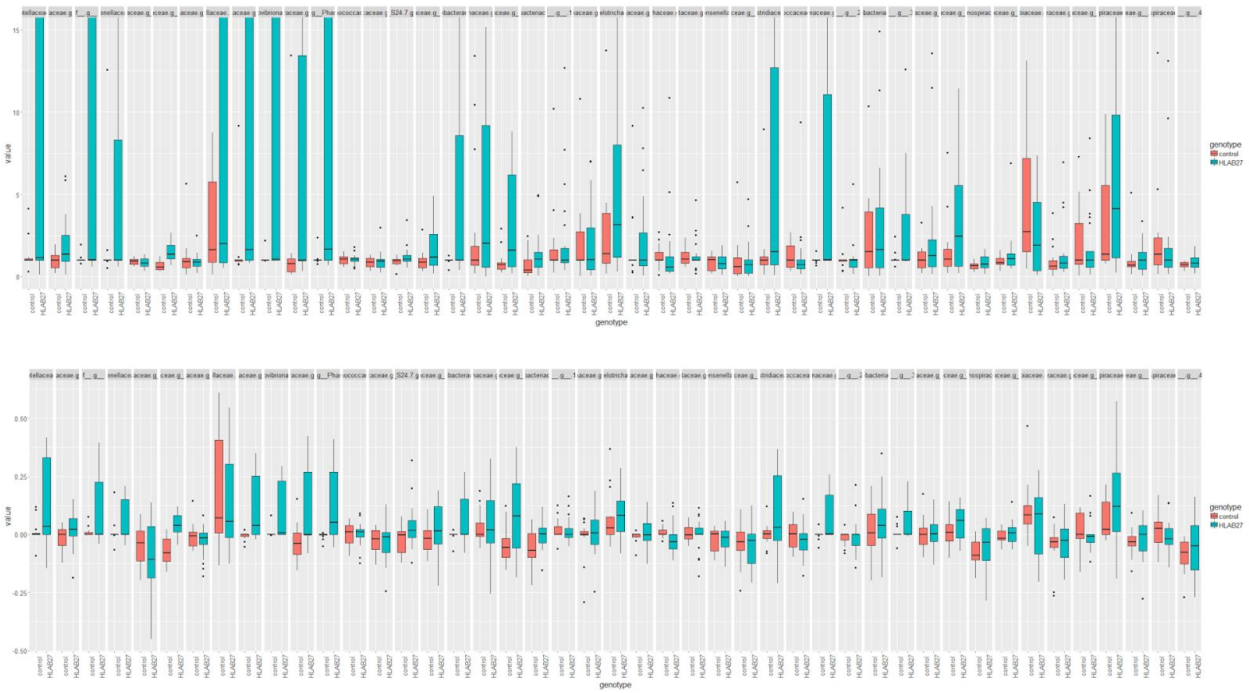


Figure 7. Boxplot of ICI (A) and IgA Index (B) showing the distinct IgA response repertoire of the transgenic HLA-B27 rat from the control rat. This illustrates the fact that no matter how we quantify the IgA response there is a distinct difference between our control and HLA-B27 rats. Horizontal bars represent group medians. N = 32 controls 40 HLA-B27 animals per group. Outliers are shown as black dots.

3.3 Microbial Repertoire of the Intestinal IgA Response Intensifies in 16 Week HLA-B27 Rats

The results of our Wilcoxon signed-rank test, which looks at the IgA response of each genotype separately, identified many bacterial genera in the 16week HLA-B27 rats that exhibited a significantly non-zero IgA index or in other words were enriched in either the IgA positive or IgA negative fraction (Fig 8). The results of our non-parametric Wilcoxon rank-sum test identified several bacteria in the HLA-B27 rats that exhibited a significantly different IgA quantification metric compared to that observed in WT controls (Fig 9). To reiterate, the Wilcoxon rank-sum is a between group comparison to identify whether specific microbial genera exhibited differences in IgA coating between WT and HLA-B27 animals. This provides further evidence for our hypothesis that HLA-B27 expression alters the microbial repertoire of the intestinal IgA response by identifying bacterial genera that potentially play a role in the immunopathogenesis of disease in these 16 week old HLA-B27 rats.

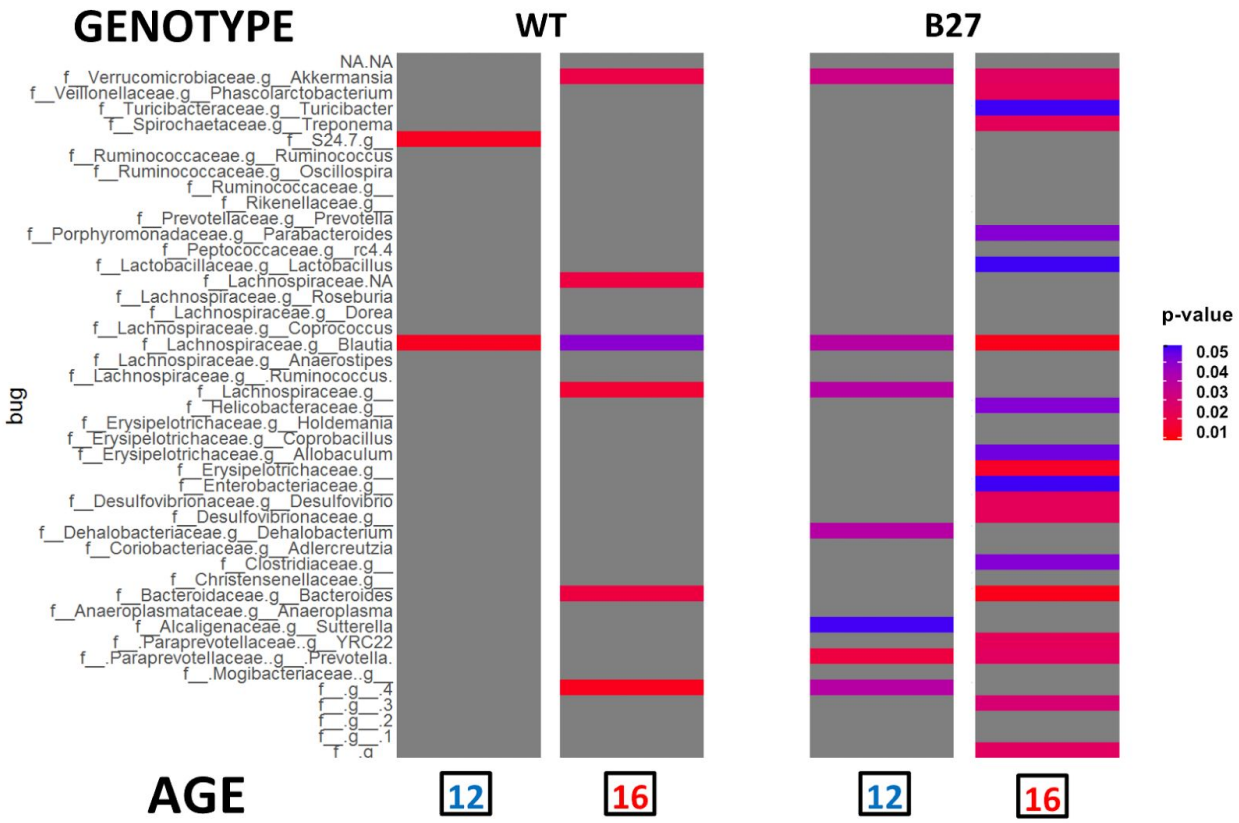


Figure 8: Wilcoxon Signed Rank test with ICI values. This figure visualizes p-values resulting from Wilcoxon Signed Rank test performed on ICI values within WT and HLA-B27 rat samples. We look at each genotype, control (WT) and HLA-B27, and see which bacterial genera are determined to be significant within each group. We further divide the control (WT) and HLA-B27 genotype groups into our two age groups, 12 and 16 weeks, to see if there is a difference in which bacterial genera are significant within each group. We show that the HLA-B27 16 week old diseased rats have more significant results according to the Wilcoxon Signed rank test. *P-values are corrected for multiple comparisons using Benjamini-Hochberg (fdr).

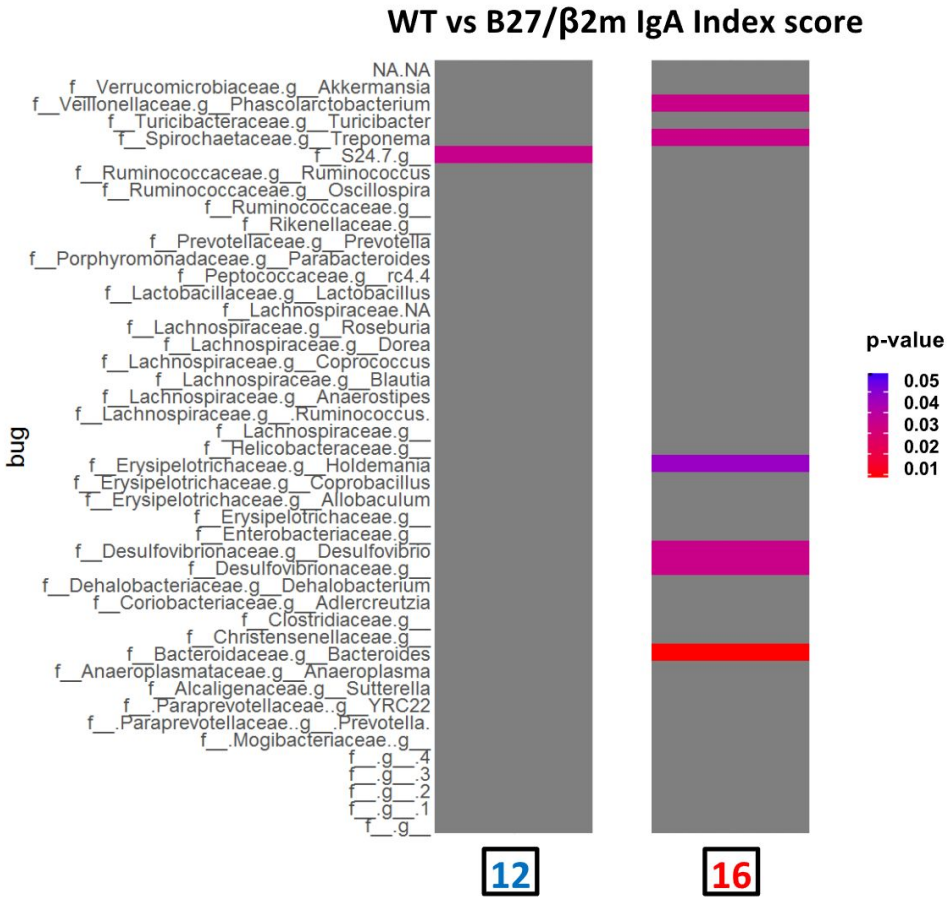


Figure 9. Wilcoxon Rank sum test with IgA Index values. The right side visualizes p-values resulting from Wilcoxon Rank Sum test comparing IgA Index values of WT vs HLA-B27 rats. We look at each genotype, control (WT) and HLA-B27, and see which bacterial genera are determined to be significant according to the Wilcoxon rank-sum group comparison. We show our 12 and 16 week groups to visualize the difference in results for each age group. We show that the HLA-B27 16 week old diseased rats have more significant results. *P-values are corrected for multiple comparisons using Benjamini-Hochberg (fdr).

We also visualized the IgA Index for HLA-B27 and control rats using a bubble plot.[3,5] A bubble plot allows us to incorporate the magnitude, sign, and significance of the IgA Index for each bacterial genera and is therefore a superior visualization to the ICI heatmap (Fig. 6). Figure 10 shows the IgA response as quantified by the IgA Index. Again, we see a distinct set of bacterial genera that are being highly IgA coated, not only in the 16 week HLA-B27 rats (Fig. 10), but also in the whole dataset as well.

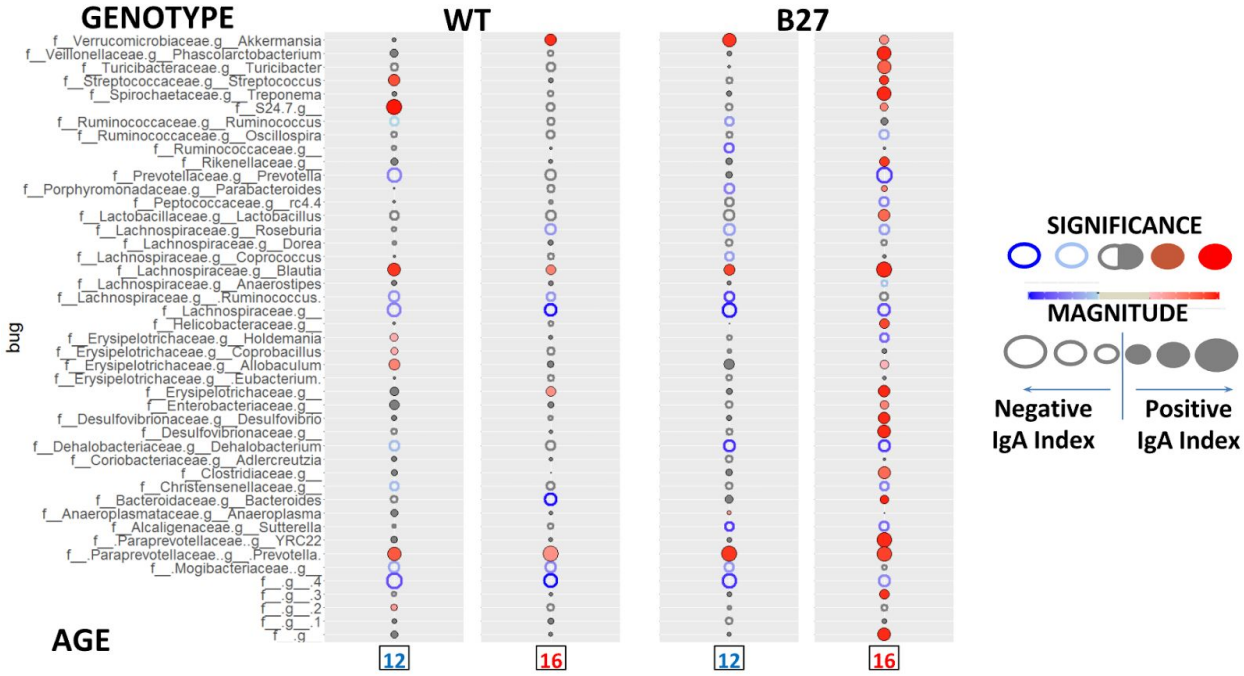


Figure 10. Bubble plots illustrating the IgA response as quantified by the IgA Index showing genotype and age groups of the HLA-B27 rats. The size of the circle represents the magnitude of the average IgA Index for the bacterial genera. The color of the circle represents a positive or negative IgA Index value, red is a positive value (IgA-coated) and blue is a negative value (not IgA-coated). The intensity of the color in the circle represents the level of significance as determined by Wilcoxon Signed Rank test within WT and HLA-B27 samples. We show that the HLA-B27 16 week old diseased rats have more significant results according to the Wilcoxon Signed rank test. *P-values are corrected for multiple comparisons using Benjamini-Hochberg (fdr).

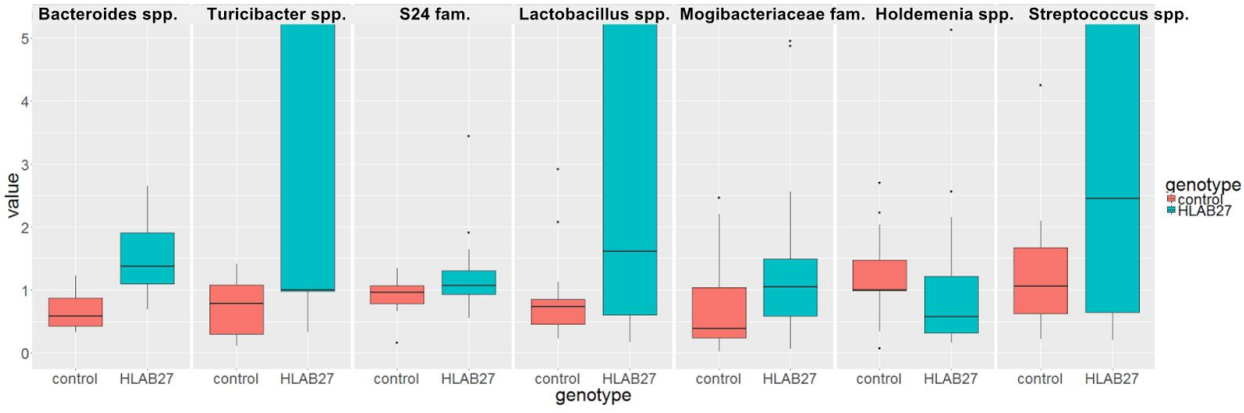
3.4 The Predictive Model Using the IgA-Index Outperformed the ICI Score Model in the HLA-B27 Rat Dataset

The variable importance plots and model evaluation metrics from our predictive modelling approach are shown in Figure 11. Variable importance plots shows us which bacterial genera best discriminate between control and HLA-B27 rat, giving us crucial insight the immunopathogenesis of AS by using predictive modelling techniques. Figure 11 shows the results of how well each of the random forest models, one built with the ICI score and the other built with IgA Index, was able to discriminate control vs. HLA-B27 rats. We show a common model evaluation metric, i.e., AUC (Fig. 11).[26]

Data set / IgA Metric	RF AUC	LDA AUC	CART AUC	Logit AUC
HLA-B27 Rat Data / ICI	0.78	0.53	0.72	0.53
16 week rats / ICI	0.95	0.61	0.78	0.53
HLA-B27 Rat Data / IgA Index	0.81	0.70	0.68	0.64
16 week rats / IgA Index	0.90	0.85	0.75	0.40

Figure 11. We evaluate the performance of our various predictive models using AUC. We notice that IgA Index is slightly more predictive than the ICI score in some cases and in some cases the ICI score is more predictive. Here we show the model evaluation results using the full HLA-B27 rat dataset and the 16 week old diseased subset. Both random forest performed ~0.80 AUC which is evidence for a predictive IgA-response signature in the HLA-B27 rats.

We built 100 different random forest models and considered the bacterial genera/features that were consistently in the top 5 most important variables to the random forest models. Focusing on our 16week old subset again we look at Figure 12. We see the bugs that were important in the ICI model are slightly different than those bugs important in the IgA Index model. The most predictive genus in both the ICI and IgA Index predictive models is **Bacteroides spp.**, however, we see several other predictive genera that are interesting as well, i.e., **Turicibacter spp**., **Lactobacillus spp**., **Streptococcus** spp., **Blautia spp., and Treponema spp.** (Fig. 12). We see again that no matter how we quantify the IgA response, there is a predictive signature/phenotype which discriminates our control and HLA-B27 rats. The AUC and ROC also inform us that there is definitely a predictive IgA signature, for certain bacterial genera, that can discriminate control and HLA-B27 rats. Both random forest performed ~0.80 AUC which is evidence for a predictive IgA-response signature in the HLA-B27 rats.



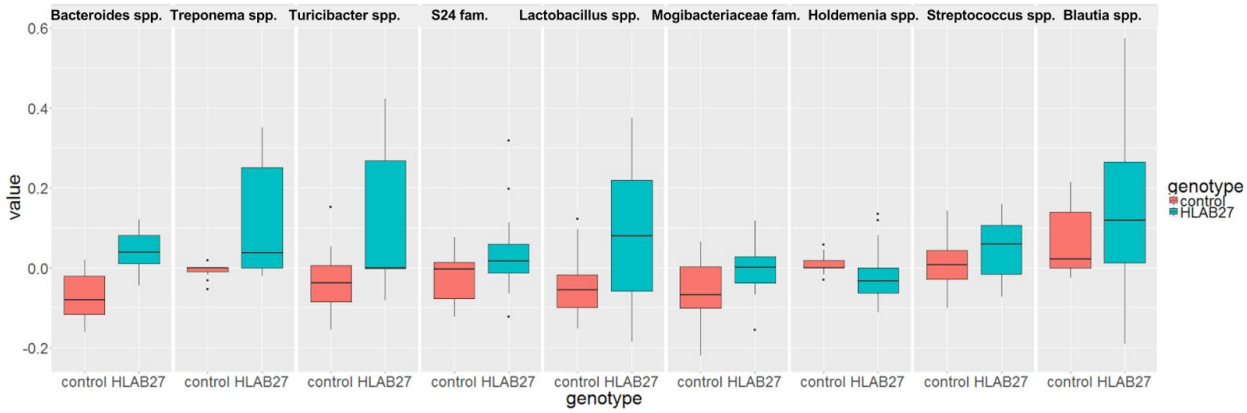


Figure 12. Discriminative Immunogenic Bacteria. Shown here are the bacterial genera that best discriminate 16 week old HLA-B27 and control rats. (Top) and (Bottom) show the important features/bacterial genera for the ICI score predictive model and IgA Index predictive model respectively. These bacteria provide us with a list of potential inflammatory 'pathobionts' that could be driving disease-driving.

3.5 Characterized the IgA-response to Intestinal Microbiota in HLA-B27 Humans

Even before running the IgA-SEQ workflow (FIg. 4) on the HLA-B27 human dataset we were aware of the heterogeneous nature of human gut microbiota. It has been shown that the composition of the human gut-microbiome can fluctuate drastically from individual to individual. This individual variation/heterogeneity has also been seen in IgA-SEQ studies.[2] Palm et. al. mentioned that they found a unique signature of IgA-coating for each individual human in their study further stating that the ability to identify 'a unique combination of IgA-inducing inflammatory commensals' illustrates the advantages of IgA-SEQ over conventional community-wide 16S profiling.[2] IgA-SEQ gives us an understanding of the immunopathogenesis of a disease and allows us to identify 'inflammatory commensals'

Here we present the results of running the HLA-B27 human dataset through the IgA-SEQ workflow (Fig. 4). These results represent some of the first insights into the inflammatory bacterial taxa that are potentially driving AS in our human subjects (supplemental materials).

3.6 Predictive Signature More Elusive in Heterogeneous HLA-B27 Humans

Our predictive modeling approach did not perform as well in the HLA-B27 human data due to the heterogeneity and individualized nature of each human's inflammatory IgA-coated gut microbiota phenotype (Fig, 13). However, we can still gain insight into which bacterial genera / predictive features are contributing to the model using the same methodology mentioned in section **2.4**. Some of the top contributing bacterial genera's IgA Index value distributions are shown in Figure 18 these include: a bacteria from family **Ruminococcaceae,** family **Mogibacteriaceae, Escherichia spp. , Bifidobacterium spp. , Odoribacter spp. .**

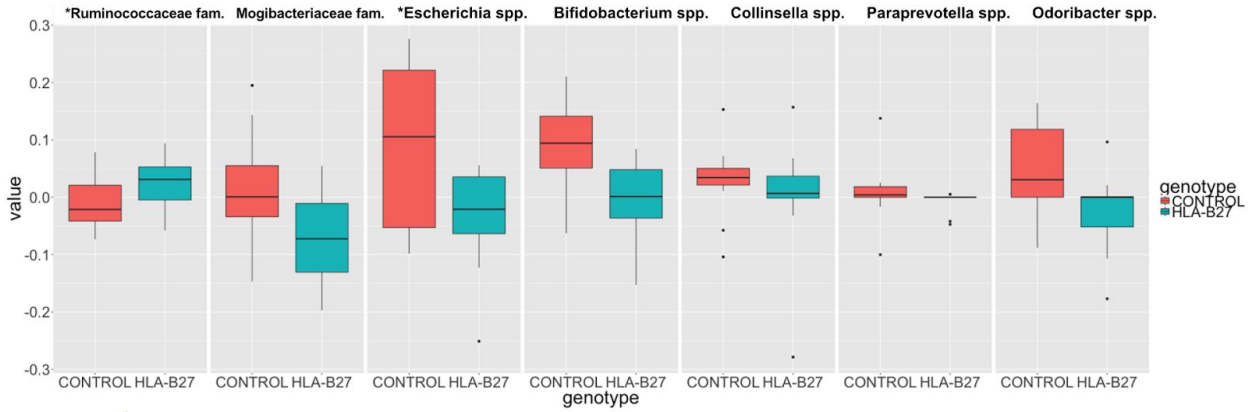


Figure 13. Shown here are the bacterial genera that best discriminate HLA-B27 and control humans based on their IgA Index score and contribution to the random forest model. We visualize these bacterial genera in boxplots to get a sense of the overall distribution of IgA Index values for control and HLA-B27 humans. The important features/bacterial genera for the IgA Index predictive model are shown.These bacteria provide us with a list of potential inflammatory 'pathobionts' that could be driving disease-driving. *Ruminococcaceae and Escherichia are the two most interesting genera.

3.7 Several Highly IgA-coated Bacterial Genera in the HLA-B27 Rats Exhibit Similar IgA-coating in the HLA-B27 Humans

In our human HLA-B27 dataset we found that some of the highly IgA coated bacteria in the HLA-B27 rats were also exhibiting similar IgA coating in the HLA-B27 humans. Specifically, **Lactobacillus spp.** and **Blautia spp.**, displayed similar IgA coating according to the IgA Index (Fig.14). This is a very interesting result and gives merit to the HLA-B27 rat model's translational potential. The fact that these two bacterial genera are being targeted by the immune system in both the human and rats is encouraging for future studies. These future studies could possibly involve isolating, culturing, and mono-associating these potentially immunogenic microbes intro germ free mice.

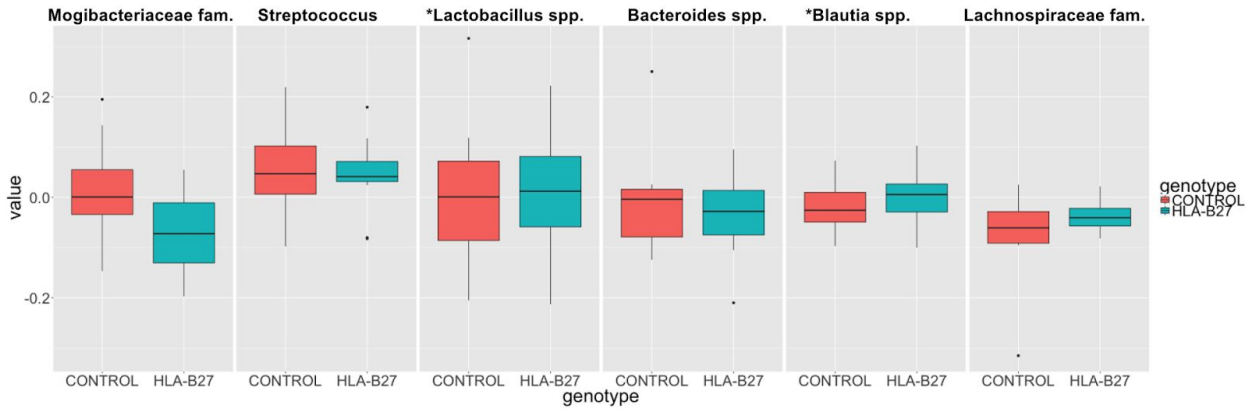


Figure 14. Shown here are the distribution of IgA Index values for the control and HLA-B27 human bacterial genera that were found to be most discriminative in the HLA-B27 rat data set. We visualize these bacterial genera in boxplots to get a sense of the overall distribution of IgA Index values for control and HLA-B27 humans. *Specifically, **Lactobacillus spp**. and **Blautia spp**., displayed similar IgA coating

according to the IgA Index metric. This is a very interesting result and gives merit to the HLA-B27 rat model's translational potential.

Discussion

4.1 HLA-B27 expression alters the microbial repertoire of the intestinal IgA response

Here we have presented the results from the examination of the intestinal immune response to the gut-microbiota in our HLA-B27 rat model and HLA-B27 humans. We used the novel IgA-SEQ technique to test the hypothesis that HLA-B27 expression alters the microbial repertoire of the intestinal IgA response. IgA-SEQ and our analysis revealed HLA-B27 expression dramatically altered the targeting of gut microbes by the intestinal IgA response. We have shown previously the IgA response to gut microbes is strongly elevated in the HLA-B27 rat model, and now we have elucidated the specificity of this response. We have also provided a first glimpse of what bacteria might be contributing to the immunopathogenesis of AS in our smaller HLA-B27 human dataset.

We found that the IgA response specifically targeted a subset of bacterial genera identified as potentially immunogenic in the HLA-B27 rat data set. This group of potential 'pathobionts' or potentially disease driving intestinal bacteria were candidates to investigate further in our smaller human dataset of fecal samples from HLA-B27 positive and negative individuals (n=12-12/group). We observed that two bacterial genera, **Lactobacillus spp.** and **Blautia spp.**, were targeted by IgA in both HLA-B27 rat models and in HLA-B27 humans. These are potential bacterial taxa to isolate and study further.

We found that the difference in the IgA response between 12 and 16 week old transgenic HLA-B27 rats was interesting in that it gave us some insight into how the IgA response of these animals develops with age. The subset of bacteria that progressively become more immunogenic, or more highly IgA coated, as these rats age could act as a first step in elucidating the immunopathogenesis of AS by creating a microbial biomarker for early development of AS in humans. Translating results from these findings into our smaller human dataset will be more difficult because we only have a limited number of samples and the age difference is nominal in the human cohort.

4.2 Overlap Between IgA-coated Microbes in AS, CD, and SpA

When we compare the gut microbes that have been targeted by the IgA response in our HLA-B27 rat model to other disease models, such as IBD and CD, we see some similarities. We expect to see similarities due to the fact that there is so much genetic, clinical, and micro-biological overlap between IBD and AS.[19] The micro-biological overlap in the IgA response between IBD and AS is apparent when we look at a recent study focused on IBD/Crohn's disease.[2] Not only is there overlap between the bacterial genera being investigated in these studies and ours, but several bacterial genera exhibit a similar IgA response. For example, the bacterial genus, **Lactobacillus spp.** (Fig.12), which was found to be significantly IgA enriched in Palm et. al was also found to be significantly IgA enriched in the HLA-B27 rat dataset and we see similar level of IgA-coating in the HLA-B27 human dataset as well.[2] What is extremely interesting about this overlap in the IgA response of the same genus in two independent studies, is that, this provides novel insight into how the host IgA response fits into the already known genetic, clinical, and micro-biological overlap between IBD and AS.[19]

When we compare the gut microbes that are highly IgA-coated in the HLA-B27 human dataset to other human IgA-SEQ studies, we see some interesting similarities. Palm et. al. included Blautia as on of the microbes in their IgA+ consortia culture.[2] We also found that **Blautia spp.** was more highly IgA-coated in our HLA-B27 humans and HLA-B27 rats. Viladomiu et. al. concluded that 'IgA-coated Escherichia coli' was more highly coated in patients with Crohn's Disease (CD) associated Peripheral spondyloarthritis (SpA) compared to patients with only CD.[18] Interestingly, we find **Escherichia spp.** to be highly coated in our healthy controls compared to our HLA-B27 human subjects. While this might be surprising in light of the CD-peripheral SpA findings, it should be noted that patients in this specific study were exclusively HLA-B27 negative to avoid potential overlap with other forms of B27-dependent SpA. A plausible explanation may be the the individual variation/heterogeneity of the IgA-response to gut microbiota we see in each of our human HLA-B27 subjects and the relatively low 'n' of each study. This also indicates however that whereas a robust IgA response to Escheria spp. may be a pathophysiological mechanism in HLA-B27 negative individuals that develop SpA, different mechanisms (at least in individuals without disease as in our study) may operate prior to development of disease in the 90% of individuals with SpA that are HLA-B27 positive. An interesting shared feature of both of our studies however is the enrichment of IgA-coated members of the **Ruminococceae** family in HLA-B27+ individuals and patients with IBD (irrespective of the presence/absence of peripheral spondyloarthritis). In this latter study the identified microbial OTU was the Ruminococceae family member *Eubacterium coprostanoligenes*.[18]

4.3 Predictive Modeling Provides Insight into HLA-B27 Rat and Human Datasets

Our predictive modeling approach revealed that the predictive models built using the IgA Index were able to discriminate genotype slightly more accurately than the predictive models built using the ICI score in some cases and in others the ICI score was slightly more accurate. We see again that no matter how we quantify the IgA response, there is a predictive signature/phenotype which discriminates our control and HLA-B27 rats. The AUC and ROC also inform us that there is definitely a predictive IgA signature, for certain bacterial genera, that can discriminate control and HLA-B27 rats. Both random forest models performed ~0.80 AUC which is evidence for a predictive IgA-response signature in the HLA-B27 rats. The most predictive genus in both the ICI and IgA Index predictive models is **bacteroides spp.**, however, we see several other predictive genera that are interesting as well, i.e., **Turicibacter spp**., **Lactobacillus spp**., **Streptococcus** spp., **Blautia spp., and Treponema spp.** (Fig. 12).

The predictive modeling on the HLA-B27 human data did not perform nearly as well but still provided insight into certain bugs that may be interesting to keep an eye on until we have a larger human sample. When we consider the individual variation/heterogeneity that has also been seen in other IgA-SEQ studies our results confirm that each HLA-B27 human harbors a unique IgA-targeted inflammatory microbial repertoire.[2] With a larger human sample we may see some general trends appear like we did in the HLA-B27 rats, and our predictive model may perform better with the larger sample size.

Aside from these results, there is an obvious bias in the ICI score, which is, the artificial inflation by substituting the small substitution value and the fact that the metric is unbounded and can blow up to infinity. However, the ICI score substitution makes sense in terms of maximizing the metric's ability to detect any taxa that were only found in the positive fraction and not in the negative. The IgA Index is more unbiasedly designed in that it is 'log-normalized', i.e., the values are bounded and range from negative one to positive one. Combining the predictive modelling results with our intuitive observations about the two metrics, we recommend that the IgA Index becomes the standard IgA response quantification metric for any experiments that use IgA-SEQ.

Future Work

5.1 Possible Predictive Intestinal Immune-targeted Bacterial Biomarker

The results from this experiment have translational potential. The most interesting result is two microbes being targeted by the intestinal IgA response in the HLA-B27 rat, also show up as highly IgA coated in the human data set. These two microbes, **Lactobacillus spp. and Blautia spp.**, are candidates to study further. One option is to isolate these two bacterial strains and mono-associate them with germ free mice (Fig. 15).

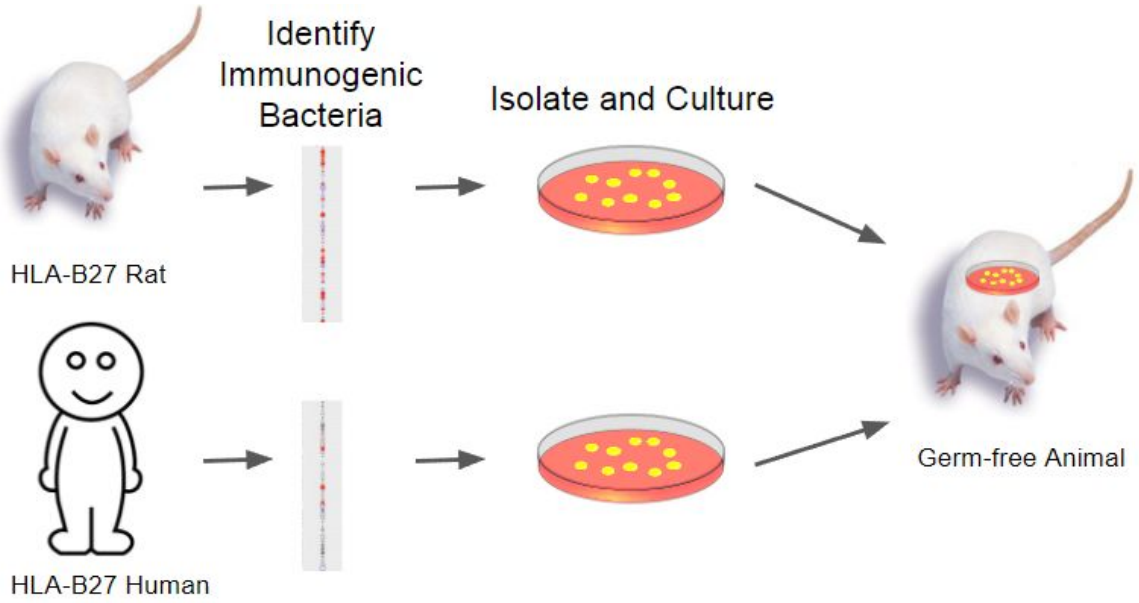


Figure 15. Outline of future work for research. This visualizations shows how we can identify potentially immunogenic microbes in our data sets, rat and human, then culture and isolate those strains. We can then mono-associate these potentially immunogenic bacteria into a germ free animal to see if these bacteria induce disease, i.e., bowel and joint inflammation.

Imagine a prognostic or predictive intestinal immune-targeted bacterial biomarker that could predict early onset of AS or monitor progression of disease. It's possible to identify those immunopathogenic microbes that are possibly driving disease development in an individualized manner like we have shown. The IgA-SEQ workflow gives us crucial insight into which bacterial taxa need to be targeted, eliminated, and possibly replaced with another bacterial taxa. An immune-targeted bacterial biomarker for AS would be the long term goal of this study. However,

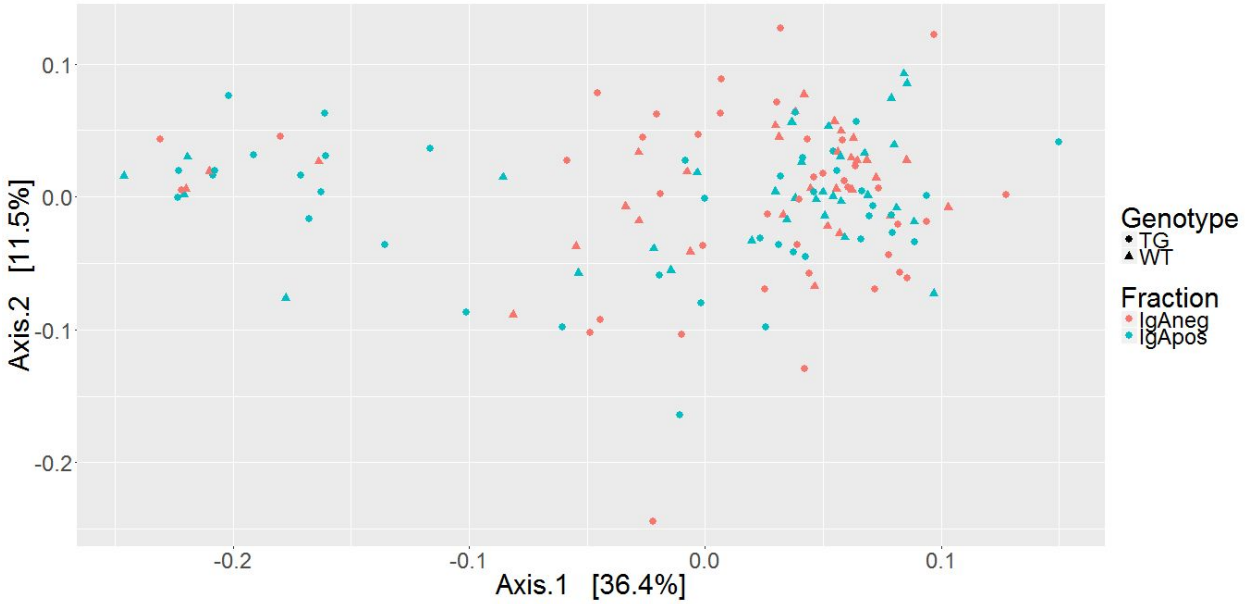
the IgA-SEQ workflow will hopefully allow for even more IgA-SEQ data sets to be analyzed including a study on age related macular, a study on urgency urinary incontinence/overactive bladder syndrome, and a study on pre-Arthritis.

Works Cited

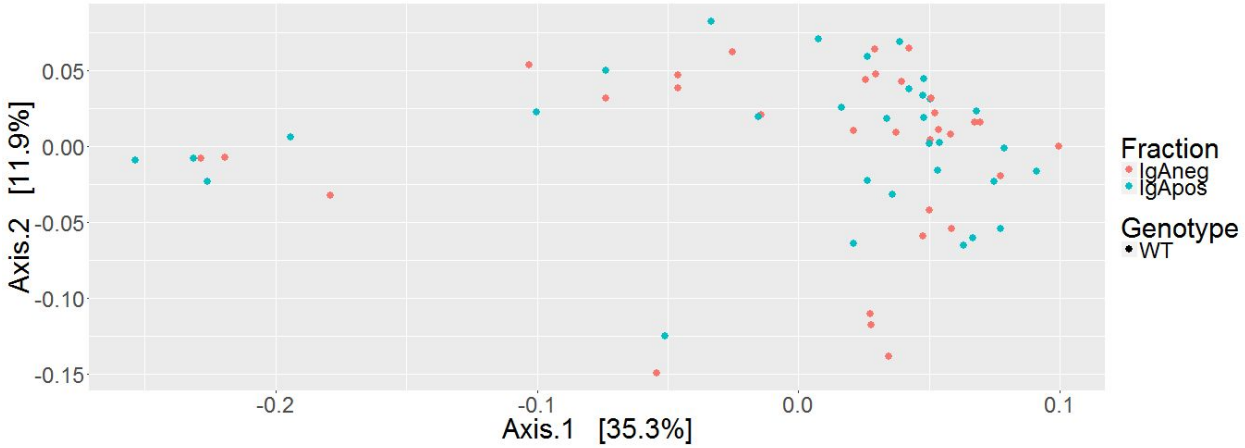
1. M. J. Asquith, P. Stauffer, S. Davin, C. Mitchell, P. Lin, and J. T. Rosenbaum. Perturbed Mucosal Immunity and Dysbiosis Accompany Clinical Disease in a Rat Model of Spondyloarthritis. Arthritis Rheumatol., 2016 Sep;68(9):2151-62.
2. N. W. Palm, M. R. De Zoete, T. W. Cullen, N. A. Barry, J. Stefanowski, L. Hao, P. H. Degnan, J. Hu, I. Peter, W. Zhang, E. Ruggiero, J. H. Cho, A. L. Goodman, and R. A. Flavell. Immunoglobulin A coating identifies colitogenic bacteria in inflammatory bowel disease. Cell. 2014 Aug 28;158(5):1000-1010.
3. J. D. Planer, Y. Peng, A. L. Kau, L. V Blanton, I. Malick Ndao, P. I. Tarr, B. B. Warner, and J. I. Gordon. Development of the gut microbiota and mucosal IgA responses in twins and gnotobiotic mice. Nature. 2016 Jun 9;534(7606):263-6..
4. K. Honda and D. R. Littman. The microbiota in adaptive immune homeostasis and disease. Nature. 2016 Jul 7;535(7610):75-84.
5. A. L. Kau, J. D. Planer, J. Liu, S. Rao, T. Yatsunenko, J. I. Gordon et.al. Functional characterization of IgA-targeted bacterial taxa from malnourished Malawian children that produce diet-dependent enteropathy. HHS Public Access. Sci Transl Med. 2015 Feb 25;7(276):276ra24.
6. Flavell, Richard, Noah PALM, ZOETE Marcel DE, and Yale University. Compositions and Methods for Identifying Secretory Antibody-bound Microbes. Patent WO2014159510 A1. 2 Oct. 2014. URL: https://www.google.com/patents/WO2014159510A1?cl=en.
7. N. Segata, J. Izard, L. Waldron, D. Gevers, L. Miropolsky, W. S. Garrett, and C. Huttenhower. Metagenomic biomarker discovery and explanation. Genome Biol. 2011 Jun 24;12(6):R60.
8. S. Mandal, W. Van Treuren, R. A. White, M. Eggesbø, R. Knight, and S. D. Peddada. Analysis of composition of microbiomes: a novel method for studying microbial composition. Microb Ecol Health Dis. 2015 May 29;26:27663.
9. Paulson JN, Stine OC, Bravo HC, Pop M. Differential abundance analysis for microbial marker-gene surveys. Nat Methods. 2013 Dec;10(12):1200-2.
10. Littman and Pamer. Role of the commensal microbiota in normal and pathogenic host immune responses. Cell Host & Microbe. 2011; 10:311-323.
11. Bunker, J. J. et al. Innate and adaptive humoral responses coat distinct commensal bacteria with immunoglobulin A. Immunity 43, 541–553 (2015).
12. J Gregory Caporaso, Justin Kuczynski, Jesse Stombaugh, Kyle Bittinger, Frederic D Bushman, Elizabeth K Costello, Noah Fierer, et. al. QIIME allows analysis of high-throughput community sequencing data. Nature Methods. 2010 May;7(5):335-6.
13. The Biological Observation Matrix (BIOM) format. http://biom-format.org/
14. Wilcoxon Rank Sum and Signed Rank Tests.R Documentation. https://stat.ethz.ch/R-manual/R-devel/library/stats/html/wilcox.test.html

15. UI Guidelines vs. Usability Testing. Microsoft Corporation. 2000. URL: https://msdn.microsoft.com/en-us/library/ms997578.aspx.
16. Fitting Generalized Linear Models. R Documentation. https://stat.ethz.ch/R-manual/R-devel/library/stats/html/glm.html
17. Ben J. Callahan,Kris Sankaran,Julia A. Fukuyama,Paul J. McMurdie , Susan P. Holmes. Bioconductor Workflow for Microbiome Data Analysis: from raw reads to community analyses [version 2; referees: 3 approved]. F1000Research 2016, 5:1492 Last updated: 05 APR 2017.
18. Viladomiu M, Kivolowitz C, Abdulhamid A, Dogan B, Victorio D, Castellanos J. IgA-coated E. coli enriched in Crohn's disease spondyloarthritis promote T H 17-dependent inflammation. Science Translational Medicine. 2017 Feb 8;9(376).
19. Asquith M, Elewaut D, Rosenbaum JT. The role of the gut and microbes in the pathogenesis of Ankylosing Spondylitis. 2014 Oct;28(5):687-702.
20. McMurdie PJ and Holmes SP. (2013) phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLoS ONE. 8(4):e61217.
21. Random Forest R package. https://cran.r-project.org/web/packages/randomForest/randomForest.pdf
22. Rowan S, Jiang S, Korem T, Szymanski J, Chang ML, Szelog J, et. al. Involvement of a gut-retina axis in protection against dietary glycemia-induced age-related macular degeneration.Proc Natl Acad Sci USA. 2017 May 30;114(22):E4472-E4481.
23. Hopfner F, Künstner A, Müller SH, Künzel S, Zeuner KE, Margraf NG, et. al..Gut Microbiota in Parkinson Disease in a northern German cohort.Brain Res. 2017 May 12. pii: S0006-8993(17)30180-4.
24. Karstens L, Asquith M, Davin S, Stauffer P, Fair D, Gregory WT, Rosenbaum JT, McWeeney SK, Nardos R. Does the Urinary Microbiome Play a Role in Urgency Urinary Incontinence and Its Severity? Front Cell Infect Microbiol. 2016 Jul 27;6:78.
25. Knights D, Costello EK, Knight R. Supervised classification of human microbiota. FEMS Microbiology Reviews. 2011 Mar;35(2):343-59.
26. ROCR R package. https://rocr.bioinf.mpi-sb.mpg.de/ .
27. SILVA. https://www.arb-silva.de/.
28. Greengenes. http://greengenes.lbl.gov/Download/.
29. Costello et al. Brief Report: Intestinal Dysbiosis in Ankylosing Spondylitis. Arthritis Rheumatol. 2015 Mar;67(3):686-691.
30. Baeten D et al. Ankylosing spondylitis and bowel disease. Best Pract Res Clin Rheumatol. 2002 Sep;16(4):537-49.
31. Taurog et al. The germfree state prevents development of gut and joint inflammatory disease in HLA-B27 transgenic rats. J Exp Med. 1994 Dec 1;180(6):2359-64.
32. Hammer et al. Spontaneous inflammatory disease in transgenic rats expressing HLA-B27 and human beta 2m: an animal model of HLA-B27-associated human disorders. Cell. 1990 Nov 30;63(5):1099-112.

Supplemental Materials

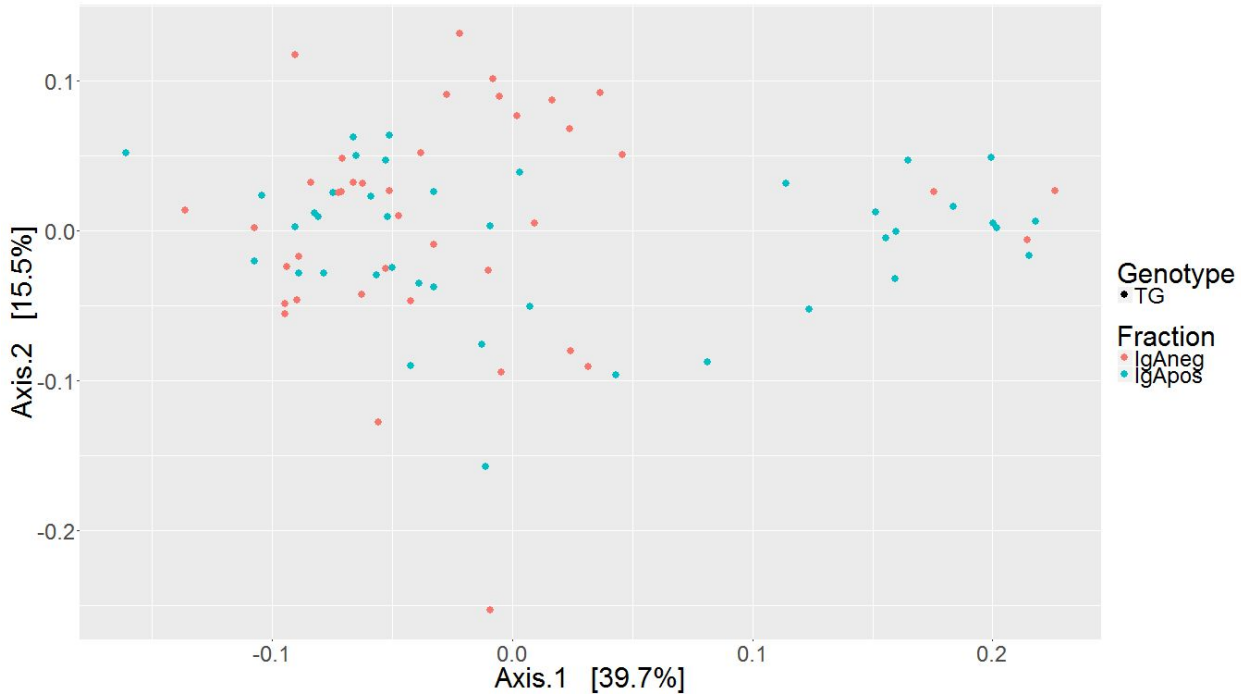


S1. Shown here is a PCoA plot of the HLA-B27 rat data set labeled with different shapes for genotype, WT or TG, and colored by the IgA-SEQ fraction, IgA-negative or IgA-positive. This PCoA plot was constructed using the weighted-unifrac distance metric. We can see some separation between the bacterial communities in our IgA-pos and IgA-neg fractions.

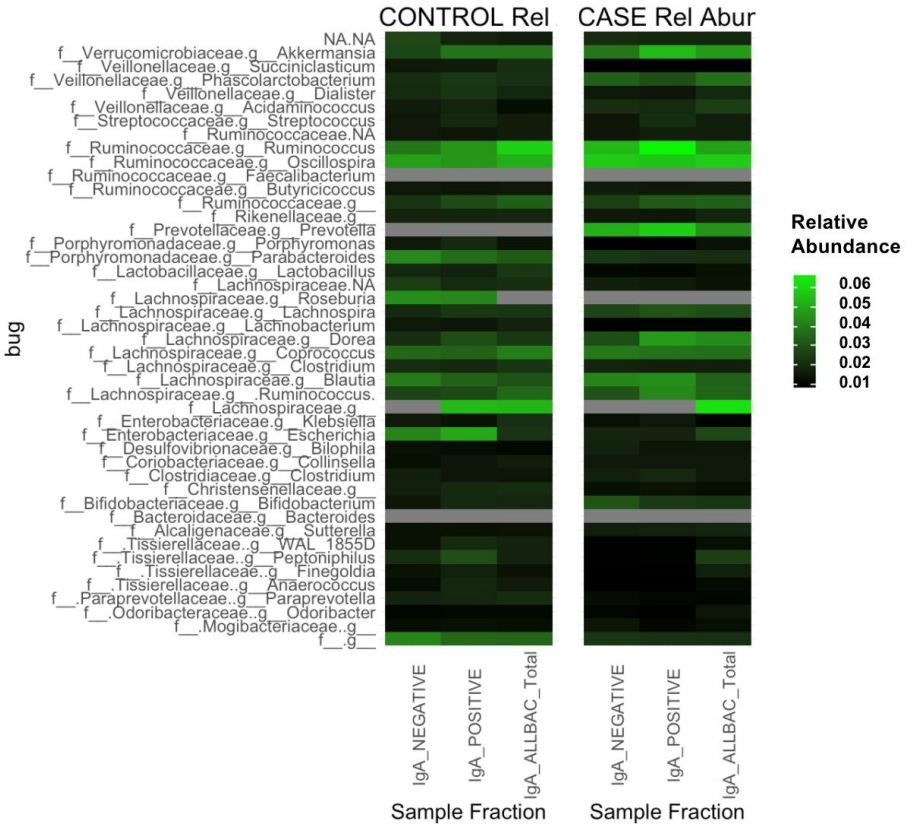


S2. Shown here is a PCoA plot of the HLA-B27 rat data set which displays the control (WT) rats and is colored by the IgA-SEQ fraction, IgA-negative or IgA-positive. This PCoA plot was constructed using the weighted-unifrac distance metric. We can see some separation between

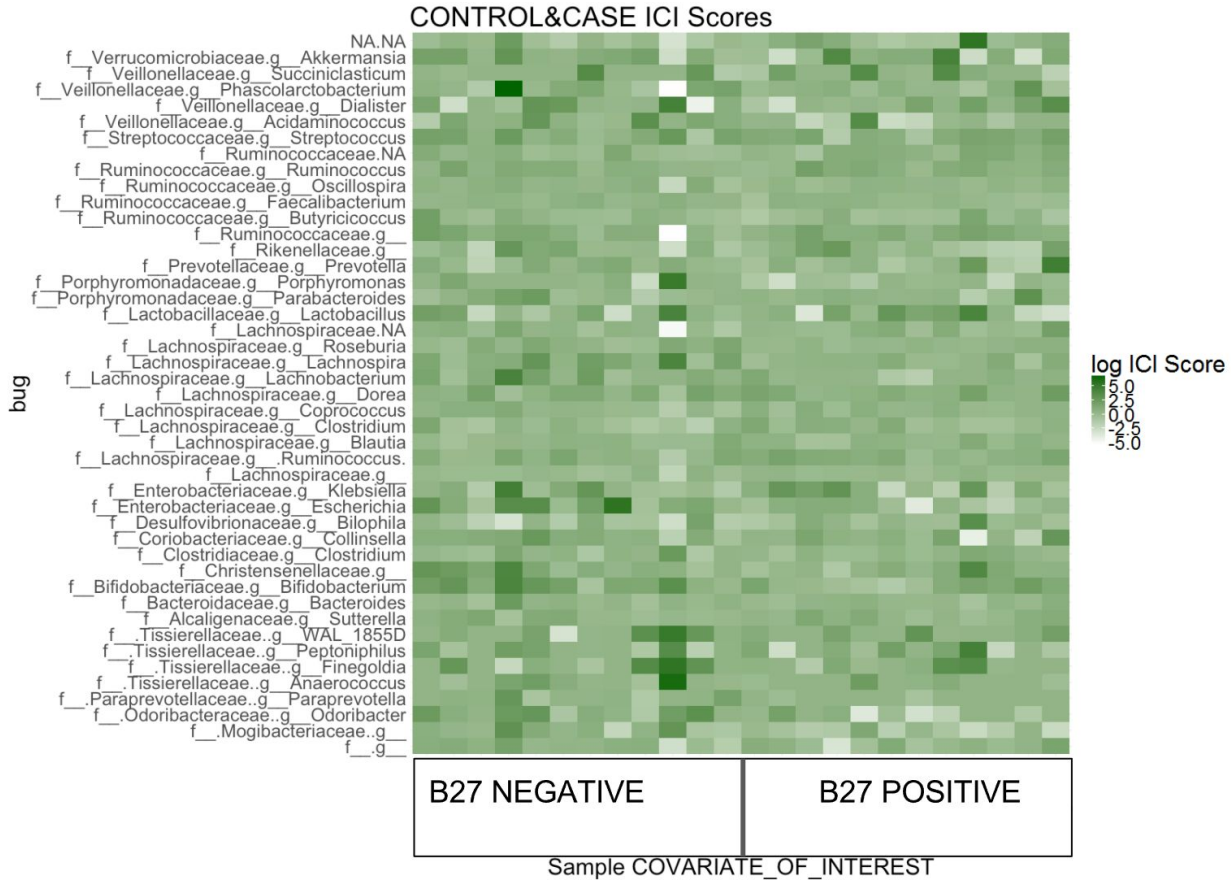
the bacterial communities in our IgA-pos and IgA-neg fractions.



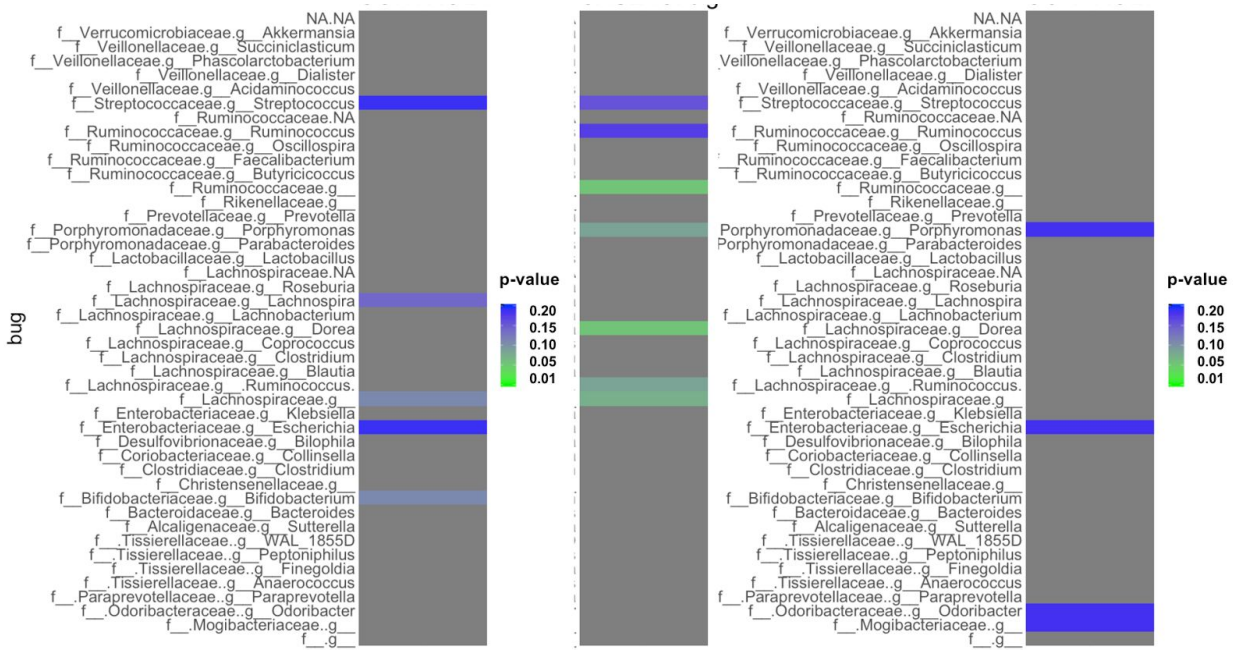
S3. Shown here is a PCoA plot of the HLA-B27 rat data set which displays the HLA-B27 (TG) rats and is colored by the IgA-SEQ fraction, IgA-negative or IgA-positive. This PCoA plot was constructed using the weighted-unifrac distance metric. We can see some separation between the bacterial communities in our IgA-pos and IgA-neg fractions.



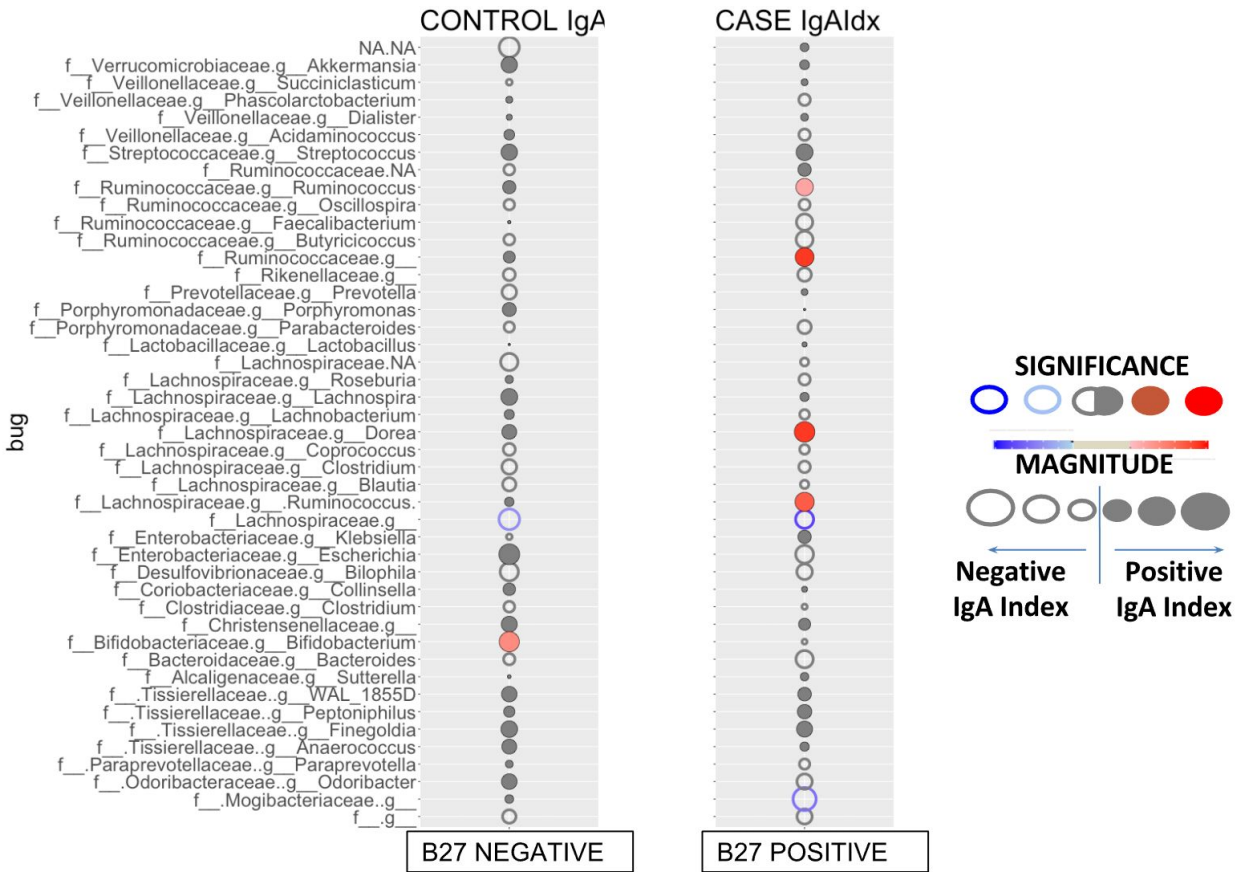
S4. Heat maps are used to illustrate the relative abundance values for the top 45 most abundant OTUs in the AllBac Fraction of the HLA-B27 Human data set (microbes that may exhibit equivalent colonization in two populations yet elicit a different host response). The average relative abundance values for IgA-, IgA+ and AllBac fractions are shown for control and CASE (HLA-B27) humans. We notice some bacterial genera that are more abundant in the IgA-positive fraction of the HLA-B27 humans.



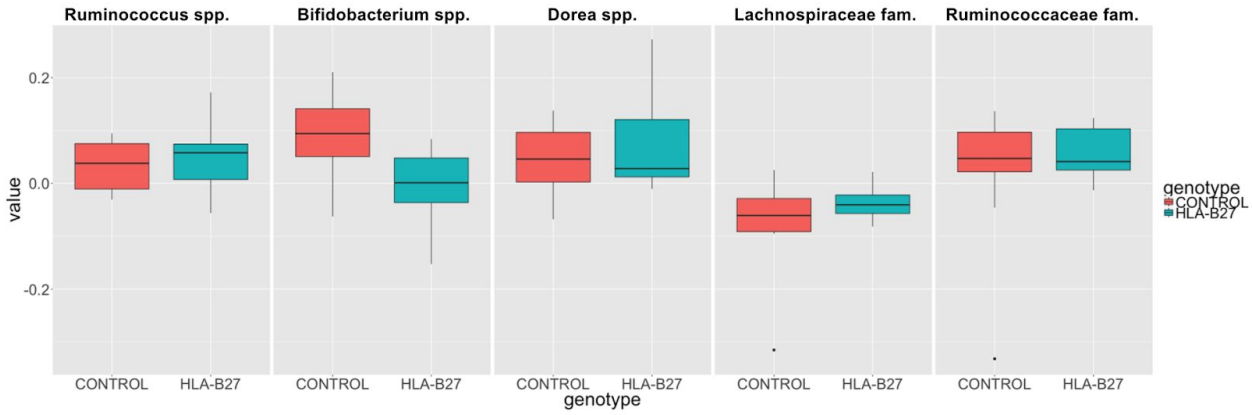
S5. Heat map illustrating the distribution of ICI scores for the top 45 most abundant OTUs in the AllBac Fraction of the HLA-B27 Human data set (microbes that may exhibit equivalent colonization in two populations yet elicit a different host response). The control human subjects are shown on the left and the HLA-B27 humans on the right. The IgA response as quantified by the ICI score identifies several OTUs that are 'highly IgA coated' in the HLA-B27 humans and the control humans. This plot illustrates the unique signature of IgA-coating for each individual human we see in our data set. *log ICI score is shown here for better visualization.



S6. This figure visualizes p-values resulting from Wilcoxon Signed Rank test performed on ICI values within control and HLA-B27 human samples. We look at each genotype, control and HLA-B27, and see which bacterial genera are determined to be significant within each group. *P-values are corrected for multiple comparisons using Benjamini-Hochberg (fdr). The right side visualizes p-values resulting from Wilcoxon Rank Sum test comparing ICI values of control vs HLA-B27 humans. We look at each genotype, control and HLA-B27, and see which bacterial genera are determined to be significant according to the Wilcoxon rank-sum group comparison. *P-values are corrected for multiple comparisons using Benjamini-Hochberg (fdr).



S7. Bubble plots illustrating the IgA response as quantified by the IgA Index showing genotype and age groups of the HLA-B27 Humans. The size of the circle represents the magnitude of the average IgA Index for the bacterial genera. The color of the circle represents a positive or negative IgA Index value, red is a positive value (IgA-coated) and blue is a negative value (not IgA-coated). The intensity of the color in the circle represents the level of significance as determined by Wilcoxon Signed Rank test within control and HLA-B27 human samples. We observe that there are a few bacterial genera with significantly large IgA Index values. *P-values are corrected for multiple comparisons using Benjamini-Hochberg (fdr).



S8. Shown here are the bacterial genera that were found to be significant according to the Wilcoxon signed rank test (Fig.16) in the human HLA-B27 dataset. We found 4 significant genera in the HLA-B27 positive humans and found 1 significant genus in the HLA-B27 negative humans. We see that the differences are not nearly as clear as we saw in the HLA-B27 rat dataset but we still see some interesting differences in the level of IgA coating between control and HLA-B27 humans in a few bacterial genera.